

GenCore version 5.1.4-p5-A578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 28, 2003, 12:00:02 ; Search time 259 Seconds

(without alignments)
6576.213 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455

Sequence: 1 atggggcttcctcgaata.....atatatgatgaagatttg 10628

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Attached: 301932 segs, 80129803 residues

Total number of hits satisfying chosen parameters: 603864

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame_n2p.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US10087013/runat_28042003.102848.20384/app.query.fasta_1.10823
-DB=Published Applications_AA -OFMT=fastan -SUFFIX=n2p.rapp -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -SPRAT=1 -END=1 -MATRIX=blomsu62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR=SCORE-pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10087013.@CGN_1.1.119.@runat_28042003.102848.20384
-NCPU=6 -ICPU=3 -NO_XLPEX -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications_AA:

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCFUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCFUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	2972.5	15.3	2710	US-10-153-273-12
2	1032.5	5.3	700	US-10-153-273-10
3	665.5	3.4	1086	US-09-924-154-15
4	649.5	3.3	921	US-10-153-273-8

10/087-013

5	618	3.2	1435	9	US-10-153-273-4	Sequence 4, Appl
6	607	3.1	1501	10 <td>US-09-924-154-17</td> <td>Sequence 17, Appl</td>	US-09-924-154-17	Sequence 17, Appl
7	600.5	3.1	1421	10 <td>US-09-924-154-13</td> <td>Sequence 13, Appl</td>	US-09-924-154-13	Sequence 13, Appl
8	597	3.1	1143	10 <td>US-09-924-154-14</td> <td>Sequence 14, Appl</td>	US-09-924-154-14	Sequence 14, Appl
9	431	2.2	749	9 <td>US-10-153-273-6</td> <td>Sequence 6, Appl</td>	US-10-153-273-6	Sequence 6, Appl
10	382.5	2.0	972	10 <td>US-09-924-154-16</td> <td>Sequence 16, Appl</td>	US-09-924-154-16	Sequence 16, Appl
11	365	1.9	411	9 <td>US-10-153-273-20</td> <td>Sequence 20, Appl</td>	US-10-153-273-20	Sequence 20, Appl
12	343.5	1.8	1115	9 <td>US-10-153-273-2</td> <td>Sequence 2, Appl</td>	US-10-153-273-2	Sequence 2, Appl
13	304.5	1.6	6281	10 <td>US-09-815-242-12996</td> <td>Sequence 12996, A</td>	US-09-815-242-12996	Sequence 12996, A
14	291	1.5	2125	10 <td>US-09-919-172-29</td> <td>Sequence 29, Appl</td>	US-09-919-172-29	Sequence 29, Appl
15	289	1.5	1639	9 <td>US-10-087-464-10</td> <td>Sequence 10, Appl</td>	US-10-087-464-10	Sequence 10, Appl
16	279	1.5	362	9 <td>US-10-153-273-18</td> <td>Sequence 18, Appl</td>	US-10-153-273-18	Sequence 18, Appl
17	279	1.4	2789	10 <td>US-09-801-574-57</td> <td>Sequence 57, Appl</td>	US-09-801-574-57	Sequence 57, Appl
18	276.5	1.4	2285	10 <td>US-09-932-183A-2</td> <td>Sequence 2, Appl</td>	US-09-932-183A-2	Sequence 2, Appl
19	276	1.4	411	9 <td>US-10-153-273-19</td> <td>Sequence 19, Appl</td>	US-10-153-273-19	Sequence 19, Appl
20	273.5	1.4	311	9 <td>US-10-153-273-21</td> <td>Sequence 21, Appl</td>	US-10-153-273-21	Sequence 21, Appl
21	270.5	1.4	1805	9 <td>US-09-820-843A-73</td> <td>Sequence 73, Appl</td>	US-09-820-843A-73	Sequence 73, Appl
22	249.5	1.3	2785	10 <td>US-09-801-574-8</td> <td>Sequence 8, Appl</td>	US-09-801-574-8	Sequence 8, Appl
23	243	1.2	2478	10 <td>US-09-815-242-5816</td> <td>Sequence 5816, Ap</td>	US-09-815-242-5816	Sequence 5816, Ap
24	243	1.2	2478	10 <td>US-09-815-242-12967</td> <td>Sequence 12967, A</td>	US-09-815-242-12967	Sequence 12967, A
25	232	1.2	3158	10 <td>US-09-815-242-12611</td> <td>Sequence 12611, A</td>	US-09-815-242-12611	Sequence 12611, A
26	232	1.2	2139	9 <td>US-10-023-219-4</td> <td>Sequence 4, Appl</td>	US-10-023-219-4	Sequence 4, Appl
27	228	1.2	2139	10 <td>US-09-727-384-6</td> <td>Sequence 6, Appl</td>	US-09-727-384-6	Sequence 6, Appl
28	226	1.2	2268	10 <td>US-09-815-242-5635</td> <td>Sequence 5635, Ap</td>	US-09-815-242-5635	Sequence 5635, Ap
29	226	1.2	2368	10 <td>US-09-815-242-12389</td> <td>Sequence 12389, A</td>	US-09-815-242-12389	Sequence 12389, A
30	222.5	1.1	665	9 <td>US-09-820-843A-107</td> <td>Sequence 107, Appl</td>	US-09-820-843A-107	Sequence 107, Appl
31	220	1.1	1331	9 <td>US-10-087-464-50</td> <td>Sequence 50, Appl</td>	US-10-087-464-50	Sequence 50, Appl
32	219.5	1.1	2025	10 <td>US-09-815-242-5703</td> <td>Sequence 5703, Ap</td>	US-09-815-242-5703	Sequence 5703, Ap
33	218.5	1.1	324	9 <td>US-10-153-273-17</td> <td>Sequence 17, Appl</td>	US-10-153-273-17	Sequence 17, Appl
34	217.5	1.1	5795	10 <td>US-09-815-242-12610</td> <td>Sequence 12610, A</td>	US-09-815-242-12610	Sequence 12610, A
35	214	1.1	807	9 <td>US-09-820-843A-108</td> <td>Sequence 108, Appl</td>	US-09-820-843A-108	Sequence 108, Appl
36	209	1.1	2434	10 <td>US-09-815-242-5835</td> <td>Sequence 5835, Ap</td>	US-09-815-242-5835	Sequence 5835, Ap
37	206	1.1	1400	10 <td>US-09-764-176-7</td> <td>Sequence 7, Appl</td>	US-09-764-176-7	Sequence 7, Appl
38	206	1.1	1404	10 <td>US-09-811-045A-1</td> <td>Sequence 1, Appl</td>	US-09-811-045A-1	Sequence 1, Appl
39	205	1.1	2835	9 <td>US-09-885-535-4</td> <td>Sequence 4, Appl</td>	US-09-885-535-4	Sequence 4, Appl
40	204.5	1.1	1183	9 <td>US-09-870-755-45</td> <td>Sequence 45, Appl</td>	US-09-870-755-45	Sequence 45, Appl
41	204	1.0	2437	10 <td>US-09-815-242-5834</td> <td>Sequence 5834, Ap</td>	US-09-815-242-5834	Sequence 5834, Ap
42	203	1.0	1719	9 <td>US-10-012-895-378</td> <td>Sequence 378, App</td>	US-10-012-895-378	Sequence 378, App
43	203	1.0	1719	9 <td>US-09-895-793-378</td> <td>Sequence 378, App</td>	US-09-895-793-378	Sequence 378, App
44	203	1.0	1719	9 <td>US-09-895-814-378</td> <td>Sequence 378, App</td>	US-09-895-814-378	Sequence 378, App
45	203	1.0	1719	10 <td>US-09-759-143-378</td> <td>Sequence 378, App</td>	US-09-759-143-378	Sequence 378, App

ALIGNMENTS

RESULT 1
US-10-153-273-12

Sequence 12, Application US/10153273

Patent No. US20020169305A1
GENERAL INFORMATION:

APPLICANT: Slim, Kim L.

Chitnis, Chetan

Miller, Louis H.

Peterson, David S.

Su, Xin-zhaun

Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knodbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273

FILING DATE: 21-May-2002
 CLASSIFICATION: <Unknown>
 APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 US-10-153-273-12

Alignment Scores:
 Prod. No.: 1,9e-182 Length: 2710
 Score: 2972.50 Matches: 865
 Percent Similarity: 40.06% Conservative: 442
 Best Local Similarity: 26.51% Mismatches: 873
 Query Match: 15.28% Indels: 1083
 Gaps: 120

US-10-087-013-1 (1-10628) x US-10-153-273-12 (1-2710)

25 ATTATTAATGGGATGACATCATTTAGAGGAGTCTTAAGCCCTATTATA 84
 6 LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspValIleGluAsp 22
 85 AAGAAAGTCACAAAGTCAGAAATGTTTGAACGTTTATGCCAAATTAAGCAT 144
 23 GluSerIleLysHisMetPheAspArgIleGlyLysAspValIleAsp 38
 145 CCATCAAAA---TATGCAAAAGACATGATGATTCGTGAAGGGATTTGACAAAGCA 201
 39 LysValLysGluLysGluLysGluLysGluLysGluLysGluLysGluLys 58
 202 GAATTCGTGGTCTCTTACAGCCAGTAATAAGCATATTATTATTCATCATCA 261
 59 LysPheGluLysAsnGluSerAspProGlnThrProGluAsp-----Pro 73
 262 TGTATTTAGATCATTAAGCAACATACATTAATTTAGCGTATGATGATGAATTTGAGCAT 321
 74 CysAspLeuAspHisLysIleHisThrAsnVal-----ThrThrAsnValIleAsn 90
 322 CTTGGCCATGTAGAGCAAAACCCGATTTGATAGATGAAGAAATCTGATGAT 378
 91 ProCysAlaAspArgSerAspValIleArgPheSerAspIleLysGluLysCysThrHis 110
 379 AATTAATACGTAATTAATAAGAAATAATGATCTATACCTGCGGCCACCTAGAGCA 438
 111 AsnArgIleLysAspSerGlnGlnIleLysAsnLysGluLysAlaAspProLysArgVal 130
 439 CGACATATGCTGATTAATAAAGCTTGAAGCTTAATGATTAATAATCCCAAAATATTAT 498
 131 LeuHisValCysAspGlnAsnLeuGlnIleGluProLysIleThrAsnThrHis 150
 499 GATTATTTGGAAATGCTAGTATGATTAAGCAAAATGCAAGGTAATCAATTTGTAATAT 558
 151 AsnLeuLeuValAspValCysMetAlaIleLysPheGlnGluLysIleThrGlnAsp 170
 559 CATCCA---CATAAAGCACT-----TCAGACGCTGTACTGCTCTTGA 600

171 TyrProLysTyrGlnAlaThrTyrGluLysSerProSerGlnIleCysThrMetLeuAla 190
 601 CGAAGTTTTCAGATATAGTATGATTAAGCAAAATGATATGTTTAAACCAAGTGC 660
 191 ArgSerPheAlaAspIleGlyLysPheValIleArgGlyAspLeuTyrLeuGlnLysPro 210
 661 CATGAC-----AAAGTAGAAACGGGCTCCGAGAGGTTTCAAGAAATA 705
 211 GlnGluIleLysGlnArgGlnGlnLeuGlnAsnLeuLysThrIlePheGlyLysIle 230
 706 CATGAT-----GGAATGGAAGATGAAGTAAATAATGATTAATCTGATGATCT 756
 231 TyrGluLysLeuAsnGlyLysGlnAlaIleArgTyrGluLysAsp-----ProGlu 246
 757 GGAATTTATTAATTAATTAAGCAAGACATGCTGGAATGATTAATTAATTAAGTGGAA 816
 247 -----PhePheLysLeuArgGluAspIlePheThrAlaAsnArgGluThrValIlePhe 264
 817 GGTATTAATGATGATGATCATCATTAATTAATCTGATTAATTAATGCAAGAAATATCA 876
 265 AlaIleThrCysAsnAla---TrpGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 282
 877 CCATTAATTTCAATTCCTAATGCGGCATTAACCAAGAAAGTTCTCCATTAATTAAT 936
 283 GlyLysArgThrLysGlyTyrCysLysCysAsnAspAspIleValProThrTyrPheAsp 302
 937 TATGTCCTCATATTTATACGTTGTTGACGAATGGGAGAAAGATTTGCCCAAAAGA 996
 303 TyrValProGlnIleLysAspValIleArgTyrPheGlnGluLysPheCysArgLysLys 322
 997 AATTAATTAATTAAGTCAAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1044
 323 AsnLysLysIleLysAspValIleArgAsnCysArgGlyLysAspLysGluLysAsp 342
 1045 CGCTATATTTGTGCTAATTAAGCAATGATTTGACCAACTATTGGAAGAAAGTAT 1104
 343 Arg---TyrCysSerArgAsnGlyLysAspCysGluLysThrLysArgAlaIleLys 361
 1105 TTGCATTTGATTAATTAAGTCAAGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1164
 362 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProLysValAsp 381
 1165 TTGGGATTAACCAAGCACTTTAAATAAAGAAAGAAATTAATGAAAGAAATTAACA 1224
 382 IleAsnAsnGlnLysGlnIlePheAspLysGlnLysLysTyrAspLysGlnIleLys 401
 1225 TCATAT-----TATGCAACGATTAACAAATTTGCTCAAT 1257
 402 LysTyrGluAsnGlyAlaSerGlyLysSerArgGlnLysArgAspAlaGlyThrThr 421
 1258 AATTAATTAATGCAAA---TATTAATTAACAAATTTTGAAGAACTTAAGAAACCAATAT 1314
 422 ThrThrAsnTyrAspGlyTyrGlnLysLysPheTyrAspGlnLeuAsnLysSerGluTyr 441
 1315 GCAACTAATGACACTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1362
 442 ArgThrValAspLysPheLeuGlnLysLeuSerAsnGlnIleCysThrLysValLys 461
 1363 -----AAAGGAGCTTACGAGGAGAAAGATTAATTTACTTAACAGTCT 1410
 462 AspGlnGluLysGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 481
 1411 -----GATGACAAAGGATTAATTTATGCTTACAGAAATTTGCAAGATGCT 1455
 482 SerGlyThrAsnValGlnIleSerGlnIleThrPheTyrArgSerLysTyrCysGlnProCys 501
 1456 CCGCAGCTCGGGGCAATGATGATGATTAATAATACACACAC---AATGATGATTAATGAT 1512
 502 ProTyrCysGlyValLys-----LysValAsnAsnGlnLysLysSerAsnGln 517
 1513 CGTGAACGTGAATTAAT-----GAAGACTTAATTAACCTCATGAGGGGT 1554
 518 TrpGlnGluLysAsnAsnGlnLysCysLysSerGlyLysLeuTyrGluProLysProAsp 537

QY 1555 GTGAACCTACTAATATACCTGCTCTTATAGTGAATGACAGCTGATATTACAA 1614
DB 538 LysGluGlyThrThrIleThrIleLeuLysSerGlyLysGlyHisAspIleGlu 557
QY 1615 AATTGAAAATTTTGTAC-----AGCTCA 1641
DB 558 LysLeuAsnLysPheCysAspGluLysAsnGlyAspThrIleAsnSerGlyLysGly 577
QY 1642 ACTAATTAACAAGATTAATAAT-----CAAAATGGAATG 1680
DB 578 ThrGlyGlySerGlyGlyLysAsnSerGlyArgGlnGlnLeuThrGlnGlnPheCys 597
QY 1681 TATTATAGAGTGAATAATTAATAGATGAATGAACAAATATCTGAAATCAATAT 1740
DB 598 ---TyrLysGlyGluAspValValLysValGlyHisAspGluAspGlnGlnAspTyr 616
QY 1741 GATTAACCTAG-----ATAATA----- 1758
DB 617 GluAsnValLysAsnAlaGlyLeuGlyLeuGlyLeuLysAsnGlnLysLysGln 636
QY 1759 -----TCATTTCATTAATTT 1773
DB 637 GluGlyGlyAsnThrSerGlyLysGluProAspGluIleGlnLysThrPheAsnProPhe 656
QY 1774 TTGGAATTTGGGTACATATTATTAGAGGATACTATTAGTGAATGACAACTTAA 1833
DB 657 PheTyrTyrTrpValAlaHisMetLeuLysAspSerIleHisTrpLysLysLeuGln 676
QY 1834 ACTGTATAT-----ATAATCAACCCAGCATTCG-----ATGATGAATGTAACAAATTCG 1887
DB 677 ArgCysLeuGlnAsnGlyAsnArgIleLysCysGlyAsnAsnLysCysAsnAspCys 696
QY 1888 TTAATGTTTTCACAGATGGTTAAACAAAGAAAGAAATGCAATGATTAAGAAACATG 1947
DB 697 GluCysPheLysArgTrpIleThrGlnLysLysAspGlnTrpGlyLysIleValGlnHis 716
QY 1948 TTCACAAAAGAAAGATATACG----- 1971
DB 717 Phe---LysThrGlnAsnLysLysGlyArgGlySerAspAsnThrAlaGluLeuIle 735
QY 1972 -----CAATCGTATATGATATATTAATTAATCTTTTGAAGGATATTT----- 2016
DB 736 ProPheAspHisAspTyrValLeuGlnTyrAsnLeuGlnGluLysLeuLysGlyAsp 755
QY 2017 -----TTTAAAGTTATGATTAACCTGACAAAGATGAAGCAAAATGCG 2058
DB 756 SerGluAspAlaSerGluGlnLysSerGluAsnSerLeuAsnAlaGluGlnAla----- 773
QY 2059 AAGAACTTATGAAAATATATAAAGAAAAAATGAGTTTCCATTTGGAAATAT 2118
DB 774 -----GlnGluLeuLysHisLeuArgGluIleIleGlnSerGluAspAsnAsp 789
QY 2119 AGGAC-----TATTAGAGATGCATACCTGTTTA 2154
DB 790 GlnGluAlaSerValGlyLysValThrGlnGlnLysAsnIleMetAspLysLeuLeu 809
QY 2155 GATCACTTAAGAAAGATCCACAGATATG----- 2184
DB 810 AsnTyrGluLysAspGlnLysAspLeuGlyLeuGlnIleHisGlnAspGlnGlnGln 829
QY 2185 ---AAGACAAATATATCAACAGACATGTGAACATCCCATATATGCAACAAACCG 2241
DB 830 LysGluLysGlyAspLysAsnGlnCysIleGlnGlnGlnLysAsnPheArgTyrAsnPro 849
QY 2242 TGTGTTAAACCTGCGAGCAGCA-----CCCACTAAATATATAAAGAAATAGCAAA 2298
DB 850 CysSerGlyGlnSerGlyAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGln 869
QY 2299 TACTTTAAAGAGTCATACGAGAGCAAGCAAAATCGTGCTTCAATTAATTAAGAA 2358
DB 870 MetHisIleLysAlaLysThrGlnLeuAlaSerArgAlaGlyArgSerAlaLeuArgGly 889

QY 2359 AAGCACACGAGAGTTATATAAAGTGGGGTAGAGAAAGAGACTTCAAGCAATTA 2418
DB 890 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlnIle 909
QY 2419 TGTAGATATATGATTAACCTTAATCGTAATCTGCTTTTCAATGACAGATGAT 2478
DB 910 CysLysIleAsnGlnLysAsnTyrSerAsnAspSerArgGlyAsnSerGlyGlyProCysThr 929
QY 2479 GGCAAAAGCAGCAGTATGCTATACAAACAAAGATTTCTGTAGAGAACTGAATGGAAGT 2538
DB 930 GlyLys---AspLysAsnHisGlyLysValArgMetArgIleGlyThrGlnTrpSerAsn 948
QY 2539 GATCCGGAACATCGCTAAAGATCACGAAGATGTTATTTGCTCTTAGAAGACAT 2598
DB 949 IleGluGlyLysLysGlnThrSerTyrLysAsnValPheLeuProArgArgGlnHis 968
QY 2599 ATATGTATACCAATTTGGAACATTTACAAAGGATGATCACCCACTTAATGTAATAT 2658
DB 969 MetCysThrSerAsnLeuGlnAsnLeuAspVal-----GlySerVal 982
QY 2659 GTT---GATGATTTAGTTAATAATCTTTTGGGGATGCTCTGATCAGCAAAATAT 2715
DB 983 ThrLysAsnAspLysAlaSerHisSerLeuGlyAspValGlnLeuAlaIleLysThr 1002
QY 2716 GAAGCAACAGATATATACATGTTAAGAAAGATTAACCTAAGGCCCAAGAA 2775
DB 1003 AspAlaAlaGluIleIleLysArgTyrLysAspGlnAsnAsnIle-----Gln 1018
QY 2776 GTAACTGACCCA-----AAGACACACAACTATCTGTGAGTATACGTTACGT 2826
DB 1019 LeuThrAspProIleGlnGlnLysAspGlnGlnAlaIleCysArgAlaValAlaGlySer 1038
QY 2827 TTTGCAATATAGCTATATTAATTCGAGAAAGATCTCGAAAGAAACGCTGACATG 2886
DB 1039 PheAlaAspLeuGlnLysPheIleIleArgGlyArgAspMetTrpAspLysSerSer 1058
QY 2887 GTAAAGCTGCAGACATTTGGAACCTTTTGGTATATACATTAAGTCACTCAAGC 2946
DB 1059 ThrAspMetGlnThrArgLeuIleThrValPheLysAsnIleLysGlnLysHisAspGly 1078
QY 2947 ---AAGGAAATGATTAATTAATGATGAT-----GCCCCAAATTTTAAATG 2994
DB 1079 IleLysAspAsnProLysTyrThrGlyAspGlnSerLysLysProAlaTyrLysLysLeu 1098
QY 2995 AGGAAATTTGTGGGAAGCTAATAGACCCAAAGTATGGGACCATGAATGCTGATATA 3054
DB 1099 ArgAlaAspTrpTrpGlnAlaAsnArgHisGlnValTrpArgAlaMetLysCys----- 1116
QY 3055 AATATTTGAGATTAATGCGACACCAATCAACCAAGTATTTGCGATATAGT 3114
DB 1117 -----AlaThrLysGlyIleIleLys----- 1123
QY 3115 GATCATACACCATGATGATATATCCACAAATTAAGATGAGTACGAAATGGCA 3174
DB 1124 ProGlnMetProValAspAspTyrIleProGlnArgLeuArgTrpMetGlnTrpAla 1143
QY 3175 GAATGTTACTGCAAGGTGCAAAAGAGACTATGATTAAGTTGAAGAGATGTAAGAG 3234
DB 1144 GluTrpTyrCysLysAlaGlnSerGlnGlnLysAspLysLeuLysLysIleCysAlaAsp 1163
QY 3235 TGTAGATTAAGATTAATGCTCAAGCTGTACGAAGAGAGTGTACAGTTGTACAGAG 3294
DB 1164 CysMetSerLysGlyAspGlyLys---CysThr---GlnLysAspValAspCysGlyLys 1181
QY 3295 TGCACAGAACTGTAATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 3354
DB 1182 CysLysAlaAlaCysAspLysTyrLysGlnGlnIleGlnLysTrpAsnGlnGlnTrpArg 1201
QY 3355 ATATATACATTAATTAACAAAGATTAATGATTAATGATTAATGATTAATGATTAAT 3414
DB 1202 LysIleSerAspLysTyrAsnLeuLeuTyrLeuGlnAlaLysThrThrAsnPro 1221
QY 3415 GGTATTGAGCTTCAGTACTGCCAAAATCATATAGAC-----AGAAATGTTATT 3465

QY 5563 ACTGCTATCAGTTTCTTCGATGCTTTGCCGAATGGGTGAAGATTTTTCGAACATAAA 5622
Db 1701 ----- 1701
QY 5623 GAAAAGAAATGGAGAAATTTGGTAGGGCGGTGAATGATTTACTTGTGTGATATGAA 5682
Db 1701 ----- 1701
QY 5683 GATAAAAGAAAGATGTACAGATGCGGTGACCAATATATAAAAATTTATAGTAGTGG 5742
Db 1702 ----- 1704
QY 5743 AAACCACTATGAAAAAACAAATCAAAAAATATGTGACAAATTAAGACAAAATATATGCC 5802
Db 1704 ----- 1704
QY 5803 GAGCATCTGTGGCAAAAGATGACAGACGCTCGCAATATTTAGACAAACAAATTTAAA 5862
Db 1704 ----- 1704
QY 5863 AAAATTTGAAAAATTAAGAGAGATTTGTGAATATATAGTATGAAGATGTGCCACA 5922
Db 1705 ----- 1706
QY 5923 CAGCGATTAAGTGTATAGTAAATATGCCCCGATCATATGACATGACCCAAA 5982
Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAGTAAATTTGTCAAGTGCACAGAGTCCACAGCGTACGAAAGGAA 6042
Db 1707 ----- 1717
QY 6043 ACACCGTACACACGGGTACTGATATCAAAAACGACGCGCATGAAAAAGAGCGAAA 6102
Db 1717 ----- 1717
QY 6103 ACAGCGCGCGCTACAAACAGCGGAAAAAGTGAATATCTAACACAGAAATGCGAGCA 6162
Db 1718 ----- 1725
QY 6163 CAACAACGAAACCGACAGACAGACACAAACAGAAACATCAACAGCAACACA 6222
Db 1726 ----- 1730
QY 6223 ACAGAAATCTGACGTGGCAACATGTAAGGCCATTCTTCGAATTAACACAGATACAG 6282
Db 1731 ----- 1747
QY 6283 GGTGGAATGAGGGTGTATATCCAAAGATGACAAATAT-----CCTAAATGGGGT 6336
Db 1748 ----- 1766
QY 6337 TGTATT-----GTAGGTAACTCTAAAGAAAAAT 6363
Db 1767 ----- 1786
QY 6364 GAAATATGATATGATGCTCTCTAGAGAAAAAATATGATTAATATATATATATAT 6423
Db 1787 ----- 1805
QY 6424 TTAATATTAAGAACT-----GAAATTAAGCTGAC 6453
Db 1806 ----- 1824
QY 6454 AATGATATAAAGAGCTTTTATTAATGCAAGCAATGAAACTCAATTTGTGCTTA 6513
Db 1825 ----- 1844
QY 6514 AATATATATATGAA-----AATCTGACAGCAAAATGAA----- 6549
Db 1845 ----- 1864

QY 6550 -----TTGCAAAATGGAACA 6564
Db 1865 ----- 1884
QY 6565 ATTCAGATGAATTTTAAAGAAATATGATATATATATATATATATATATATATATAT 6624
Db 1885 ----- 1904
QY 6625 TTTGGAAGTATTTCTATATGATTAATAAATAATTAAGTAAACAAATATGTAACAC 6684
Db 1905 ----- 1924
QY 6685 ATTCGCAATGAAATATATTAAGAAAAACAGATTAATAAATAAATGAAATATATCT 6744
Db 1925 ----- 1944
QY 6745 ATA----- 6747
Db 1945 ----- 1964
QY 6748 -----TTTGGAGAAAAATTAATAATTTATTTGGAGAAATGATATATATATAT 6804
Db 1965 ----- 1984
QY 6805 -----CATCTCACAGCAAGCAAAACGAAAAATTAACA 6840
Db 1985 ----- 2004
QY 6841 GATTAATTAAC----- 6849
Db 2005 ----- 2024
QY 6850 -----CAGTACAAAT-----GACATGACCAAACTG----- 6873
Db 2025 ----- 2044
QY 6874 -----ACGCCCTCCCTGAAGAGCTTTGTAATAAAGCCCC 6906
Db 2045 ----- 2064
QY 6907 CAATTTTGAATGATTCACAGATGCGCAAGAAATTTGTATTAAGAGAAAGAAAG 6966
Db 2065 ----- 2084
QY 6967 TGTAAAAATTTGAGGCGGCTGTAAAGAAATATGAG----- 7002
Db 2085 ----- 2104
QY 7003 -----TGTAAAT-----GCTACTAAT 7017
Db 2105 ----- 2124
QY 7018 GACGATAGACACAGAA-----TGTGAGAGCGCTGTATACATATCA 7062
Db 2125 ----- 2144
QY 7063 AATTTATTAAGAGTGAAGAACTGAATATGAAGCAAAAGAAAGTTCAAAAAGAT 7122
Db 2145 ----- 2164
QY 7123 AAAGAT----- 7128
Db 2165 ----- 2184
QY 7129 -----GGCAAAAGATTAAGATTAAT-----CCTTCTACTGAAAGAGACATA 7170
Db 2185 ----- 2204
QY 7171 GACAGGCAACATGCTCTATGATATTAACATGAAATTTAAAGAAATATATGTCGAAT 7230
Db 2205 ----- 2219
QY 7231 AAGGATGTTCTTGTATGCAAAAACCTTCTTCACACTTCACAAAACACACACACATCA 7290

```

Db 2220 LysAspCysAspProCysLeuLysPheSerValAsnCysLysLys----- 2234
QY 7291 CATCATCCGATGCTAATGATATGCCAGATCCGCTGGATTATGTTCTCGAAGAAATTAAAC 7350
Db 2235 -----Asp 2235
QY 7351 AAGTGTGAGTCTGCTGAACCTTTCAAAAAAGGAGCTATGATTCATACAAAAAATTACT 7410
Db 2236 GlucylAsp-----AsnSerLysGlyThrAspCysAlaGlnLys----- 2248
QY 7411 GAACCTAAATATACCTGATGATTTGCTAGAAAGACAGCATATATTATCTAAAGAACCA 7470
Db 2249 -----AsnSerLysAsp-----AlaThrAspIle 2256
QY 7471 GAAATTAATATGATATTACCTTGAGGAAAAAATTATACCTATTAGTGTACAAAGAA 7530
Db 2257 GluAsnGlyValAspSerThrVal-----LeuGluMetAlaGlyAlaSer 2270
QY 7531 AAGAAAGTAAATATGATGCTGACTAATATATACCTTCGATCCCTAAGAAACCTTATGCA 7590
Db 2271 AlaAspSerLysSerGlyPheAsnGlyAsp----- 2280
QY 7591 CCTGATTAATATATAGGAAAGAACCCCTTGCAAAAT----- 7629
Db 2281 -----GlyLeuGluAsnAlaCysArgIlyValaGlyIlePheGluGlyIle 2295
QY 7630 AGAGAAAGAAATGCTTTTAAAGTAGATATTGAATGGAATGTAACAAAATTCAAAATGTC 7689
Db 2296 ArgLysAsp-----GluThrLysCys-----ArgAsnValCysGly 2307
QY 7690 TATCAGAGCAAAAAAGAGTATGTGTACCTCCCAAGAAAGAACATATGTGCTTAAGAAAT 7749
Db 2308 TyrVal-----ValCysLysPro----- 2313
QY 7750 TTGATGGAATTTAAATTTGAAGACCTTAAGAGATGATTAATATCTCTAAAAATGTTGCT 7809
Db 2313 ----- 2313
QY 7810 CGAATCGACGAAATGAGAGAAATAGACATATATAAACTTCACTCAGAGAACGGGTGC 7869
Db 2314 -----GluAsnValAsnGlyGlu----- 2319
QY 7870 GCAATGATCCAAATATGATGATCTATGAAATATAGTTCGCTGATCGGTGACATAGTT 7929
Db 2320 -----Ala 2320
QY 7930 AGAGGAACAGATATGTTAGCAATGTGTGTTACTTACCTCCCGTAGAAATTAATATAT 7989
Db 2321 LysGlyLysHisIleIleGlnIleArgAlaLeuValLys-----ArgTyrVal 2336
QY 7990 AAGCTTTTGAATACATATATGAAATATGAGAAATATAAATAAGTAGAGAAATTAATAC 8049
Db 2337 GluTyrPhePheGluAspTyrAsnLysLysHisLysLysSerHisArgIleLysAsn 2356
QY 8050 AACGAT-----GTCAAAACGTTTCTGCTGCTGCTGAGTCGTAATAGAAAA 8097
Db 2357 GlyGluIleSerProCysIleLysAsnCysValGluLysTyrValAspGlnLysArgLys 2376
QY 8098 GATATTTTGAAGAGCAATGACCTGCAAGACACAGACAGATGCAAACTTTTACAAAAAGCA 8157
Db 2377 Glu-----TyrLysGluIleThrGluArgPheLysAsp----- 2387
QY 8158 AGATGATGATTTGAACGATATACATTAATACAGATAGTGTGACATAGAGACAT 8217
Db 2388 -----GlnTyrLysAsnAspAsnSerAspAsp 2396
QY 8218 CCACCTGTTGATGATTTATATACCTCAACGGTTTCGATGATGACGTAAGCTGTGAATAT 8277
Db 2397 AspAsnValAlaTyrSerPheLeuGluThrLeuIleProGlnIleThrAspAlaAsn----- 2414
QY 8278 TATTGTAAACACCTGATGAGAGATTTGAAAAATTTAAAAATCATGATCATCTGTAA 8337

```

```

Db 2415 ---AlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSerCysGly---CysSer 2432
QY 8338 ACATCGACAGATGCAAGATGATTATGATGAAATTAAGTGTGAACAGCTGTAAACGAGA 8397
Db 2433 AlaSer-----AlaAsnGluGlnAsnLysAsn----- 2441
QY 8398 TGTCAGAAATATATAAATTTGCTCTTAAATGCAATCTCATTCGATATACAAATCAAT 8457
Db 2442 ---GlyGluTyrLysAspAlaIle-----AspCysMetLeuLys 2453
QY 8458 AATATCAAAAGATTTGTATGACACACCAATATATACAAATCTCTACTTATGATCATGTT 8517
Db 2454 LysLeuLysAsp-----LysIleGlyGluCysGluLysLys 2465
QY 8518 CAAATTTTGTACAAAGTGTGAACCTTTTAAAGTGAATGTTCTGTGAGAGCTTTCT 8577
Db 2466 HisHis-----GlnThrSerAspThrGluCysSerAspThrProGlnPro 2480
QY 8578 GAATATCTTCATGAAACAGTAACTGTTGAATTAATTAATTAAGAAATGATGCTTCT 8637
Db 2481 GlnThrLeuGluAsp-----GluThrLeuAspAspIleGluThrGluGluAlaLys 2498
QY 8638 TCCATATATACGACATATGCTTTGCAAGAAACACCAAAAGTTATTAAGAA----- 8688
Db 2499 LysAsnMetMetProLysIleCysGluAsnValLeuLysThrAlaGlnGlnGluAspGlu 2518
QY 8689 ---GCTTGCAGTTGTACACTCTTCAAGATTCATGATGATTAATGTTCTTACCAGTCA 8745
Db 2519 GlyLysCys-----ValProAlaGluAsnSerGluInProAlaIleThrAspSer 2535
QY 8746 AACAAAGT 8754
Db 2536 GlyLysGlu 2538

RESULT 2
US-10-153-273-10
Sequence 10, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Shim, Kim L.
Chitnuls, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550

```

TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 700 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-10-153-273-10

Alignment Scores:
 Pred. No.: 3.5e-58 Length: 700
 Score: 1032.50 Matches: 247
 Percent Similarity: 50.288 Conservative: 110
 Best Local Similarity: 34.798 Mismatches: 250
 Identity Match: 5.31% Indels: 103
 Gaps: 22

US-10-087-013-1 (1-10628) x US-10-153-273-10 (1-700)

```

OY 406 AATGATGCTATGACCTGTCGCCACCTAGACAGCAGCATATGTCGATATAAACTTGA 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 5 AsnLysValGlyAlaCysAlaProGlyArgLeuHisLeuLysCysAspTyrAsnLeuGln 24
OY 466 GCTCTAATGATATAATCCCAAAATATTCATGATTTATGGGAATGACAGTATACA 525
    ..... ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 25 SerLysP-----ThrThrSerThrThrHisLysLeuLeuGlnValCysMetAla 42
OY 526 GCATAATACGAGAGTAAATCAATTTGTAATTAATCAT-----CCACATAAGAGAACT 576
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 43 AlaLysTyrGlnLysAsnSerIle-----AsnThrHisTyrThrGlnHisGlnArgThrAsn 61
OY 577 -----TCAGACGCTTGACTGCTGCTTCGACGAAAGTTTTCAGATATAGGTGAT 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 GluAspSerAlaSerGlnLeuLysThrValLeuAlaArgSerPheAlaAspIleGlyAsp 81
OY 625 ATGTATAGAGAAATATGATATGTT-----AAACCAATATGTCATGAC 666
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 82 IleValArgGlyLysAspLeuTyrLeuGlyTyrAspAsnLysGlnLysGlnArgLys 101
OY 667 AAATGAGAAACGGGTCCTCCGAGAGTTTTCAGAAAATACATGATGAAATGAAATGAA 726
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 102 LysLeuGlnGlnLysLeuLysAspIlePheLysLysIleHis-----LysAspVal 118
OY 727 GTAAAAAATGAT-----TACAAATCCTGATGGA---TCTGGAATTTATAT 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 119 MetLysThrAsnGlnValAlaGlnLysArgTyrIleAspAspAlaLysGlyLysAspPhe 138
OY 769 AAATTAAGAGAGCATGCGATGATGATGATGATGATGATGATGATGATGATGATGATG 828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 139 GlnLeuArgGlnLysAspTyrPhePheSerAsnArgGlnThrValTyrLysAlaLeuLys 158
OY 829 GATGATCATATTAATATCTGGATATTTTATGCAATCAGAAATATACACCATATTTTGA 888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 159 HisAlaProLysGlnLysAlaAsnTyrPheIleLysThrAlaCysAsnValGlyLysGly 178
OY 889 AATCTTAATTCGCGCATTAACAGAGAAAGTTCCTACCAATTTAATATATGCTCCCTCA 948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 179 AsnGlyGlnCysHisCysIleGlyLysAspValProThrLysPheAspTyrValProGln 198
OY 949 TATTTACGTTGTTGTCAGCAATGCGAGAGAGATTTTGGCGAAAAGAAATATATTAATG 1008
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 199 TyrLeuArgTyrPheGlnGlnTyrPheAlaGlnLysAspPheCysArgLysLysLysLys 218
OY 1009 AAAAAGCTACAGAGCTCTGCTGTAATGACAAAGAAAGCTTATATTTAGTCAATGGA 1068
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 219 GluAsnLeuGlnLysGlnCysArgAspTyrGlnGlnAsnLeuTyrCysSerGlyAsnGly 238
OY 1069 CATGATTTGACGACAACTATTTGGAAAAAGATATTTTGCATTGGATATATAGTCTACT 1128
  
```

```

DB 239 TyrAspCysThrLysThrIleTyrLysGlyLysLeuValIleGlyLysHisCysThr 258
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
OY 1129 GACTGTCGACATAATGCAAAAGTTTGAAGTTTGATGGGATACACAGAGCATTT 1188
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
DB 259 AsnCysSerValTyrCysArgMetTyrGlnThrTyrIleAspAsnGlnLysGlnPhe 278
OY 1189 AAAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1221
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 279 LeuLysGlnLysArgLysTyrGlnThrGlnIleSerGlyGlyLysSerGlySerPro 298
OY 1222 -----CATCATATTTTTCGACAGATACAAATTTGCAAT 1257
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 299 LysArgThrLysArgAlaAlaArgSerSerSerSerAspAspAsn----- 314
OY 1258 AATATTAATAGTGAATATATATAAACAATTTTATGAAAACCTTATAGGAAACGATATGCA 1317
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 315 -----GlyTyrGlnSerLysPheTyrGlyLysLeuLysGlnValGlyTyrGln 330
OY 1318 ACTAATGACACTTTTAAATTTACTAATTAAGAGAAAGTATTTGTAAGGA----- 1368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 331 AspValAlaLysPheLeuLysIleLeuAsnLysGlnGlyLysCysGlnLysGlnProGln 350
OY 1369 ---GATTTACACAGAGAAAAAGCATTTACTTTTACTTAACAGTCTCATGCAAAAGGATA 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 351 ValGlyAsnGlnLysAlaAspAsnValAspPheThrAsnGlnLysTyrValLys---Thr 369
OY 1426 TTTTATGTCGACCAATATTTGTCGCCAGTGTGCCGACTGGGGGTCAAATGATGATGATA 1485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 370 PheSerArgThrGlnLysCysLysProCysProTyrCysGlyLeuGlnLysGlyLysPro 389
OY 1486 AATACACACACAAATACATATATGATGTCGAACTGTAATTAATGACATATAAAGCT 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 390 ProTyrValLysGlyLysLysPheLysThrCysGlySerAlaLysThrLysThrTyrAspPro 409
OY 1546 CCATGGGGTGTGACGCTCAATATACAGTCTTTTATATGTCGTAATGACAGGTGAT 1605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 410 Lys-----AsnIleThrAspIleProValLeuTyrProAspLysSerGlnGlnAsn 426
OY 1606 ATTACACAAAATTAATTAATTTTGTACACGCTCAACTATATACAAAGTATAAATAAT 1665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 427 IleLeuLysLysTyrLysAsnPheCysGlnLysGlyAlaPro---GlyGlyGlnIle 445
OY 1666 CAAAAATGGGAATGCTATATATAAGAT-----CAAAATATAATATAGATG 1710
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 446 LysLysTyrGlnCysTyrTyrAspGlnHisArgProSerSerLysAsnAsnAsnAsn 465
OY 1711 AAATCGGAACAAATCTGAATCATATGATGATATCTTAAGTATATCATTTTCATAT 1770
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 466 ValGlnGlyThrTyrAspLysPheThrGlnGlyLysGlnThrValLysSerTyrAsnVal 485
OY 1771 TTTTGAATTTAGGCTTACATATTTATTAAGGATACATATTAAGGATGCAAACTT 1830
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 486 PhePheTyrAspTyrValHisAspMetLeuHisAspSerValGlnTyrLysThrGlnLeu 505
OY 1831 AAAACTGTATATAATATACAAACACG-----CATGTATGATGATGATGAT 1878
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 506 SerLysCysIleLeuAsnAsnThrAsnGlyAsnThrCysArgAsnAsnAsnLysCys 525
OY 1879 AGAAATTCGTTATGTTTTCAGACATGGGTTTAAACAAAAGAAAGAAAGAAATGATATA 1938
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 526 ThrAspCysGlyCysPheGlnLysTyrValGlnLysGlnGlnIleTyrPheMetAlaIle 545
OY 1939 AAGAACTGTTACAAAAAAGAAATATACACCATCGATATATGTAATATATATAT 1998
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 546 LysAspHisPheGlyLysGlnThrAspIleValGlnGln-----LysGly 560
OY 1999 CTTTGAAGGTTATTTTAAAGTTATGATGATAACTTGACAAAGATGACAAAGAAATG 2058
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 561 LeuIleValPheSerProTyrGlyValLeuAspLeuVal-----LeuLysGly 576
OY 2059 AAAGACTTATGCAAAATATATAAAGAAAAAAGAAATGAGTTTCCATTTGGAAAAATAT 2118
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

OY	6487	GCAATGAAACATCATTTTGTGGTTAAATATTAATTGAATAATCTCGAGCAGAAT	6546
Db	102	LysLeuGIuSerAsnIleuValGlnHisLys-----Asn	113
OY	6547	GAATTGCCAAATGACAACATTCACGATGAAATTTAAAAGAAATTAATATACATATGGT	6606
Db	114	AsnGIuTyfAsnSerLysLeuCysAspSphilLeArg-----TrpSerPheIeu	129
OY	6607	GATTATTAAGAAATATTTTTTGGAACGCAATGATTTCTAATGATTAATAAAAAATTTAACTGTA	6666
Db	130	AspTyfGIuAspIleIleIleGlyArgAspLeuIeIyLyfAsnAsnThrAspTyfIle	149
OY	6667	ACAATAGTGTACACACCATTCTCAATGAA-----AATATAGAAAAAACAGATATAA	6720
Db	150	LysGIuGInPheLyfLysIlePheAsnAsnGIuTyfAsnAsnAsnGIuIleuAsnAspGIu	169
OY	6721	-----AAAAAGATGAAGAAATTAACGTAAATATTTGGAGAA	6755
Db	170	LeuAsnAsnGIuLeuAsnAspGIuLysAsnIleLysLeuArgLysGIuTrpIleLys	189
OY	6760	AATTAATAATTTATTGGGAAGAAATGATATATGATTAACATTCATCTCCACACAGAA	6819
Db	190	TyrLysGIuAspIleTrpGIuGluMet-----	198
OY	6820	AACGAAAAAGAAAAATTAGAGATAATTACCGTACAAATGACACAGCAAATGCAGCCT	6879
Db	199	-----ThLysGIuHisAsnsp-----LysPheIleGIu	208
OY	6880	TCCCTTGAAAGATTGTNAAA--AGCCCCAATTTTGGATGGTTCACAGATGGGCA	6936
Db	209	LysCysLysTyfPheAlaLysAspGIuProGInIleValArgTrpIleGIuGIuTrpSer	228
OY	6937	GAAAGATTTTGTATTAAGAGAAAGAAACACTGTTTAAATTTGGAGCGGGCTGTAAAGAA	6996
Db	223	LysGIuPheLeuAspGIuLysAsnTyfMetLeuPheThrLeuArgAsnThrTyfAsnGIu	248
OY	6997	TATGAGTCAATGGTAGTACAGCGGTAAAGACACAAGATGTGCAGAGCGCGGTATAACA	7056
Db	249	MetAsnIleIleHisGIuAsnAsn-----CysLysGIu	259
OY	7057	TATCAAAATTTTATTAAGAAAGTGAAGTGAATATGAAAGCAAAAGAAAGTCAAA	7116
Db	260	TyrAsnLysTrpValGlnAsnArgLysLysGIuTrpThrPheLeuSerAsnGIuPhasn	279
OY	7117	AAGCATAAAGATGGCAAAAAG-----TATPAAGATTAATCCT	7152
Db	280	LysIlePheProGIuArgAsnValGlnIleHisIleSerAsnIlePheLyfGIuTyfLys	299
OY	7153	TCTACTGAAAGACATARGAAGAAAGCAACATGTGCTCATGATATTTAACATGAAATTA	7212
Db	300	GlnAsnAsnValAspIleIlePheGIuThrLeuAsnTyfGIuTyfAsnAsnPhe-----	317
OY	7213	AAAGAAATTATGTGGCATTAAGATGTGTTTGTATGACAAAACCTTCTTCACACTACA	7272
Db	318	-----CysLysGIuLysPro-----	322
OY	7273	AAAAACACACAAATCACATATCCATGCTAATGATATGCCAAGATGCTGATTAAT	7332
Db	322	-----	322
OY	7333	GTTCCTGAAGAAATTTAAACAGTGTGAGTGCTGCTGAACCTTCAAAAAAGGATCATGATT	7392
Db	323	-----GluLeuValSerAlaAlaLys-----	329
OY	7393	CATTCAAAAAAATTACTGACCTTAATAATACCTATGAATGTGTAGAGAAAAAGCATAT	7452
Db	330	TyrAsnLeuLysIleProAsnAlaLysSerPro-----ArgIle	342
OY	7453	TATTATCATAAAGACAGCAAAATAATATAGATATTACCTGGAAGAAAAATTTATACCT	7512
Db	343	TyrLysSerLysGIuIscGIuSerValPheGIuCysLysThrLysIle-----	360
OY	7513	ATTGAGTCTCAAGAAAAAGAAAGTAAAAATAGTGTGACTATTAATATCCTTGAGAT	7572

```

Db 360 ----- 360
QY 7573 CTAAGAAACCTTATGCACCTGATAATATATAGAGAGAAACCCTTGTGAATAAG 7632
Db 360 ----- 360
QY 7633 GAAGAAATCGTTTAAAGTAGATTAATGAATGGAATGTTTACAAAATTCAGATCTAT 7692
Db 361 ----- 361
QY 7693 CAGAGAAAAAAGAGTAGTGTACCTCCAGAGAGAAACATATGCTTAAAGAAATTA 7752
Db 376 ----- 376
QY 7753 GATGAATTTAA-----ATTGAAGACCTTAAGGATGTAATTAATCTCTAAATG 7803
Db 396 ----- 396
QY 7804 GTTCGTGAACTGCAGAAATGAGAAATGACATATATAAAAACCTCACTCAGAGAAC 7863
Db 410 ----- 410
QY 430 GATGCGCATGATGATCAATATGTGATGATGATGATGATGATGATGATGATGATGAT 7923
Db 430 GATGCGCATGATGATCAATATGTGATGATGATGATGATGATGATGATGATGATGAT 7923
QY 7924 ATGATGAGAGACAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7980
Db 447 ----- 447
QY 7981 ----- 7981
Db 459 ----- 459
QY 8035 GGTAGAAATTAATCAACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 8094
Db 479 ----- 479
QY 8095 AAGATATTTTGAAGCAATG-----ACGTCAAGACACCA 8130
Db 495 ----- 495
QY 8131 GAAGATGCAAACTTTTGAAGAAAGAGATGATGATGATGATGATGATGATGATGAT 8190
Db 513 ----- 513
QY 8191 CAGATATAGTGTGACATAGAGACATCCACTGTTGATGATGATGATGATGATGATGAT 8250
Db 520 ----- 520
QY 8251 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8310
Db 525 ----- 525
QY 8311 TTTAAAAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8370
Db 545 ----- 545
QY 8371 AATTAAGTGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8430
Db 558 ----- 558
QY 8431 AATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8490
Db 574 ----- 574
QY 8491 ACAAATTCCTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8550
Db 592 ----- 592
QY 8551 AGGAATGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8607

Db 599 ----- 599
QY 8608 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 8667
Db 616 ----- 616
QY 8668 ACACCAAAAGTTATAAGAGCTTGCACT---TGTCACCTACTCTGTAAGAT 8718
Db 627 ----- 627
QY 8719 ----- 8719
Db 646 ----- 646
QY 8767 TTACAACTTTTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 8826
Db 664 ----- 664
QY 8827 GCATACCTGTTCTTATTAATGATGATGATGATGATGATGATGATGATGATGATGAT 8886
Db 674 ----- 674
QY 8887 AGACATTTATGACAAAGCTATCAGCATATATATATGAAAGGTGATTAAGAAAT 8946
Db 682 ----- 682
QY 8947 TTAATAAA-----AACTTCTACTCTGCTTCACTGCA 8982
Db 689 ----- 689
QY 8983 GGACAAATTTGAGTCGCAAAATTAATTAATGGAAGAGTGTGC----- 9027
Db 709 ----- 709
QY 9028 TTTGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9087
Db 729 ----- 729
QY 9088 ATGACACTTCATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9147
Db 749 ----- 749
QY 9148 GAAATTCGTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9207
Db 759 ----- 759
QY 9208 GATATTAATTTGCTACTTCAAAAGTAAACATTAAGATGATGATGATGATGATGAT 9267
Db 761 ----- 761
QY 9268 GATGAAGAAACATAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9327
Db 764 ----- 764
QY 9328 GAAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9387
Db 766 ----- 766
QY 9388 TTTGAAGCGTCAGATTAATTAAGCAACCTGATGATGATGATGATGATGATGATGAT 9447
Db 779 ----- 779
QY 9448 AGCTTGAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9507
Db 782 ----- 782
QY 9508 AAGATCAATCTTCACTGATTAATTAAGCAATTAACCATCT----- 9546
Db 783 ----- 783
QY 9547 ----- 9547
Db 803 ----- 803

```

```

QY 9574 TCAAAAGATCTCAATGCGCTTGAGTTA----- 9603
Db 823 ThrspgIuSerSerIleGluIleSerIleProSeraspGluAsnHis 842
QY 9604 AATGATATAAATGAAATGTTACAGAAACAAAATAATGAAATGAA----- 9654
Db 843 SeraspLeuSerGluAsnHisGluAspSerAsnGluIleAsn 862
QY 9655 -----TTCAAGAGAGTCTAAATAATATATCTGTTATATTTGTTGAGAT 9705
Db 863 ProProSerThrGluIleGluAsnLeuIleHis----- 874
QY 9706 GAAACACAAATAATCATGACTGATGAAATATAAAGAGAGCAAGATTGCT 9765
Db 875 -----TyrIleThrSeraspSeraspHisGlySerIleIleSerGluIleGlu 982
QY 9766 CTTAAACACTCTATTTCTTACACCCCATGATCTTTCATCAAGACCTTATTC 9825
Db 893 ProIleGluIleThrGluIleSerProIleThrAspLys----- 905
QY 9826 TCAACACATGAGTACGACATATGATCTTAAATGATATTTGAAAGTACT----- 9879
Db 906 LysThrGluIleSerIleAlaIleGlyAspLysAsnHisGluSerValIleSerIleAspIle 925
QY 9880 -----ATCTCTGTTTATTGTA 9897
Db 926 PheGluSerGluIleHisAsnSeraspAsnArgAspArgIleValIleSerIleValIle 945
QY 9898 -----TCGCGGTTAGGTTTGTGATAGACGCTTCATTTCAAGAAATAATTC 9942
Db 946 GluAspSerSerGlySerSerMetSerThrGluSerIleArgThrAspAsnLysAspIle 965
QY 9943 AAATGCTCTGTGACTGTTGGGTATACGATATCCCAAGAGAGAGATGATGCT 10002
Db 966 LysThrSerGluAspIleAlaProSerIleAsn----- 976
QY 10003 AGCTTGAAATCCAAAATAGGTACATATAGAGTGTCAATATTAAGCAAAACA 10062
Db 977 -----GlyHisGlu 979
QY 10063 TATATATATGAGAGAGATCTAGTGAATGAATGAATATATGCGACTTACT 10122
Db 980 LysIleGlySerSerIleAspAspArgIleGlySerGluAspLysSerIleIleAspLysAsp 999
QY 10123 TCCCTGTATATCTATCCGAAAGTGAAGATGATGATATATGATATATAT 10182
Db 1000 SerGluAsnPheGluAsnAsnLysSerSerHisSerAspIleLysGluSerAsp----- 1017
QY 10183 GTACCAAGTATGCTTAATATATAATGATGATAGAGTACTAGACCATCAAAAAGG 10242
Db 1018 AsnGluIleSerThrAspLysGluSerLeuThrGlu-----GluSerProLysGly 1034
QY 10243 GATATACCAAGT-----GATGATACCAAGCAAGTATGATACCA 10278
Db 1035 AspleuGluSerValSerProSerSerIleAspMetAspLeuLysProAsnLysSerSer 1054
QY 10279 CCACGTACCAAT 10290
Db 1055 ProValThrSer 1058

```

RESULT 4 US-10-153-273-8

Sequence 8, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chilnits, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Welliams, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

```

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSER: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153, 273
FILING DATE: 21-May-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-273-8
Alignment Scores:
Pred. No.: 1,35e-33 Length: 921
Score: 649.50 Matches: 250
Percent Similarity: 39.48% Conservative: 127
Best Local Similarity: 26.18% Mismatches: 280
Query Match: 3.34% Indels: 288
Gaps: 48
US-10-087-013-1 (1-10628) x US-10-153-273-8 (1-921)
QY 2842 GATATATTCGAGAGAGATCTCTGGAAGAAAGAGGATGTAAGCTGCAAGCA 2901
Db 80 AspTyrIleThrAlaSerAspIle-----GluAsnGlyIleAsnSerIle----- 94
QY 2902 CATTGGAACCTGTTTGTATATATACATTAAGTCACTCAAGCAAGCAAGATGATAAA 2961
Db 95 -----GlyAsnIleAspMetValIleSerAspLysAspAlaAsnGly 108
QY 2962 TATAT-----GATGATGCCCAATATTTAAATTTAGGCAAAATTTGG 3006
Db 109 PheAsnGlyLeuAspAlaCysGlySerIleAsnIlePheLysGlyIleArgLysGluGln 128
QY 3007 TGGGAAGCTAATAGACCAAGATGAGCAAGCATGAATGATATTAATTTGAAG 3066
Db 129 TrpLysCys-----AlaLysVal-----CysGlyLeuAspValCysGly 141
QY 3067 GATTAATGGGACACCAATCAACACAAAGTACTTATTCGGATATAGTATACACCA 3126
Db 142 LeuLysAsnGlyAsn-----GlySerIleAspLysAspGln 153
QY 3127 TTGATGATATATCCCAAAAATTAAGATGATGATGACCAATGGGACCAATGTACTGC 3186

```


ORGANISM: *Plasmodium falciparum*
SEQUENCE DESCRIPTION: SEQ ID NO: 4;
US-10-153-273-4

Alignment Scores
Pred No.

Seed: NO:	1.56e-31	Length:	14553
Score:	618.00	Matches:	339
Percent Similarity:	36.80%	Conservative:	223
Best Local Similarity:	22.20%	Mismatches:	538
Query Match:	3.18%	Indels:	428
DB:	9	Gaps:	72

US-10-153-273-4 (1-1435)

Db
20 lysalaaragsnclutyr---AspIle-----TGG 3351

3352 AATATAATTCAGATAAATACAAAGATTTCATCATGCAACGACCAAAATGTCGTAGTAAT 3411

OY 3412 AGTGGATGTGAAGCTTCACACTGCCAAAAAATTCTTTTAAATAATGCG...TTTT

C maprvalyllylvsclutlysphensngltubeasplyslsystyrgr|yasnValglnlys 54

Db 55 Thr-----|||:::AspLysLysIlePheThrPhe 62

[illegible]

3517 GGTACTAGTGAATGCTCTGTCTATTGGTACTTAACACACGATGAAATGTTGCACA 3576

TATCTCCATGATAACAGGAATTTCGTCAATTTTAACTC
-----IleasplylsamMetSerLeuileasn-----
96

97 ---LysHisAsn-AsnGlu---GluMetPheAsnAsnAsnTyrGlnSerPheLeuSerTh 114

[illegible]

3697 GGTGCGTGGCTGTAAAGTCGATCGAAACGACAGGTACAGATAAAAACGAAAAA 3756
: |||::: : :|||

3757 AAGCGGAGAAAGGATACGGGAATGTAAACCTCTGC----- 139

b
140 -----SerArgIleAsnAsnGlyArgAsnThrSerSerAsnAsnIleValLeuSe 156

156 TAsNCysARGGILuLYsARqlvsg] vMeotIustrelolpLpLrTt--- ||| :::: ::
::: ||||| ::::: ||| :::: ::

3861 TCGATATCCCGATTGGCAATCGGAAATATAAATTTAGTGAAGACCCCTCGTGTGTAT 3920
:
|||

[illegible]

```

:|||||:::|||||:|||||
181 eProspargargilegInleucysilevalnleuser-----llellely 197

```

3981 ATATCAATCAACAAGTTATTTAAAGAGCTTTCATCAATCGCAGCAGCAAAACATT 4040
197 stbtrtvttrthttwglrmbh-wcattt:::|||||:|||||:|||||:|||||

4041 TTCTCATGCTATTATTATAAAGTAAGACGTGAGGAATGAACTCGATAAGAATT 4100
|||
|||

4101 AAAAGAAGCCTAAATTTCCGCCCCCGAATTTCCTTGG
216 ----SerGlnLeuLeuLysLysAsnSpasnlYTrasnSer----- 229

230 -----LysPheCysAsnAspLeuLysAsnSerPheLeuAspTyr 242

OY 4161 TAGAGATTTTATTTGAAACAGATATATCAAAAGTCATGTCAGAGGAAGTAACCTAAA 4220
 Db 242 rglYhIleuAlaMeGclYAsnAspKheLAspRhegLy---GlyTysSerThrIysAlaGl 261
 OY 4221 AGAGCAAAATAGATTCCTTTTCAAAAATGTCACAAAATCTCCATTAAGGAAAAACA-- 4278
 Db 261 usnLysIleGlnIleValPheLysGlyAlaHISglYIleSerGluHISLysIleY 281
 OY 4279 -----CCCAAGATGGTCAGACAGATAGCTATGAGATATGGAGAGCTATGTCATG 4331
 Db 281 sAsnPhaLgIysGluTTPTrpAsnIleuPheArgGluIysLysIleTTPGluAlaMetIeu-- 300
 OY 4332 TGCAGTAGTAAATAATGGGGCAAAAAAGATGATTTTACGGAAGACTGCTTACACAA 4391
 Db 301 -----SerGluHISLysAsnAs 306
 OY 4392 CCGTCAATTTAGTACAAAAGCACACACTTTGGAGATTTGCCAAAAGCACCAGTTT 4451
 Db 306 nIleAsn---AsnCysLysAsnIleProGlnIleGluIleu-----GlnIleTh 321
 OY 4452 ACAGTGGCTAACCGAATGGTACGACACTATGCTATACAGACAAAATATTTGAAGA 4511
 Db 321 rglntPpIleLysGluTTPHISglYIleuPheLysGluArgAspAsnArgSerIysLe 341
 OY 4512 TGTGAGAAAATGTAGTCAATGACCAATTTGAAG---TGTGATACAGATGTAAATA 4568
 Db 341 uProlYserIysCysLysAsnAsnThrIeuYrGluAlaCysGluIysGluCysIleAs 361
 OY 4569 GAATCCGAGACTAGCTTAAATAT---ATGAAAAAAAAGATGGATGCTTCACAGA 4625
 Db 361 pProCysMetIysTyrArgAspTrpIleIleuSerIysPheGluTTPHISThrIeuSe 381
 OY 4626 TAAATATACAGAGTACGACGACAAAAGATTCGATAGACACACATTTGGTAAAT 4685
 Db 381 rIysGluYrGluThrGlnLysValProlYsGlu----- 392
 OY 4686 GGTTCACAGACTATATGGAACAGATGCAACAGATTTACTTGAACAGAAATTTACTGCTAG 4745
 Db 393 rProLysGluTTPHISLysIleuAsnTyrIleu---IleLysIleSerGluAs 404
 OY 4746 TTGTGGATTAAGCCTGGAAGTGCCTGTGTGTCACAAAGAAATATATCAATTTAGAAA 4805
 Db 404 nIysAsnAsp-----AlaLysValSerIleuLeuAs 415
 OY 4806 ACAGGCTTACTATGATCCGACAAACATGTCGGTGC-----ACAAAATTTATGA 4856
 Db 415 nAsnCysAspAlaGluTyrSerIysTyrCysAspCysLysHISThrThrIleuValIy 435
 OY 4857 A-----AATGAC-----GACAAATATATACATTTTCGACTAA 4889
 Db 435 sSerValIleuAsnGlyAsnAspAsnThrIleLysGluLysArgGluHISLysLeuAs 455
 OY 4890 AGATTAAGTGCAAA---GGATTAATAAGAGGCA---AACACAGTGCCTTTAAGCGCA 4943
 Db 455 pAspPheSerLysPheIleCysAspLysAsnSerValAspThrAsnThrLysValItrpGl 475
 OY 4944 AAACAAGGCTCTAATATCAATTAACCTGAAGAATTAATGATGAGATGCTTTTCC 5003
 Db 475 uCysLysAsnProTyrIleuLeuSer-----ThrLysAspAlaCysValPr 490
 OY 5004 TTTCTGTCAGCTAGTATATGTTTTCATGTCATGTCGATGCAATTAATACAGATCCAGAGT 5063
 Db 490 oProAlaArgGlnIleuLysLeuGlyAsnIleAspArgIleTyr----- 505
 OY 5064 TAAAGATGAATAGGGTGG-----CGAAAAGATGATGGAAGGCGCGCAACGGAAG 5117
 Db 506 -----AspLysAsnLeuLeuMetIleLysGluHISLysLeuAlaIleAlaIleTyrGluSe 524
 OY 5118 GTACAAATTTGGTCAATATCTACAAAGAAAAAGAAAAAGAGAAATAAATAAAGCTCGCA 5177
 Db 524 rArgIleuLysArgIysTyrIysAsnLysAspAspLysGlu----- 538

OY 5178 TGGCACAATAATCTTATAGAGTCCCGCCTGTAGTGTATGAAATATAGTTTATAGA 5237
 Db 539 -----ValCysLysIleIleAsnLysThrPheAlaAs 549
 OY 5238 TTTAAGATATATATCTAGATATGATATTTTGAAGATGAAAAACAAAAGCCAGCA 5297
 Db 549 pIleArgAspIleIleGlyGlyThrAspTyrTrpAsnAsp-----LeuSerAsnAr 566
 OY 5298 AAATTTGAAGAAAATATTTTGAACAAAATGACATGATGCGCAAGAGATAGTAGAC 5357
 Db 566 gLysLysLeuValGlyLysIleAsnThrAsnSerLysTyrValHISArgAsnLysLysAsn-- 585
 OY 5358 TACAGCAATCCCGTAGTACTGCGCAAAATTTTCTGGAACGAAATATGGAATGTGT 5417
 Db 586 -----AspLysLeuPheArgGluTTPTrpLysValIleLysLysAspVa 601
 OY 5418 GTGAAGCGCAATATATCGCGGTACAAACCTGGTAGGATGATGGAATATGTGAAATAG 5477
 Db 601 TrpAsnValIleSerTrpValPheLys----- 610
 OY 5478 TGCAGAGATGATGAAGATCTAAAAAATGTGGTCTGTACTTCAAGATGATATATCC 5537
 Db 611 -----AspLysTh 613
 OY 5538 TATGGGAAAAATCCGATGAGATGACTGCTATCACTTTCTTGATGTTTCCGCAATG 5597
 Db 613 rValCysLysGluAspAspIleGluAsnIleProGlnPhePheArgTrpPheSerGluTr 633
 OY 5598 GGGTGAAGATTTTGCACAAATAAAGAAAGAAATGGAAATTTGGTAGGGCGCTTA 5657
 Db 633 pGlyAspAspTyrCysGlnAspLysThrLysMetIleGluThrLysValGluCysLy 653
 OY 5658 TGATTAATCTTGTGTCATGATGATTAAGATTAAGAAAGAAATGTACAGATCGTGCACA 5717
 Db 653 sGluLysProCysGluAspAsp-----AsnCysLysSerIysCysAsnSe 668
 OY 5718 ATATAAAAATTTATATGTCAGTGGGAAACACAGTATGAAAACAAATCAAAAATATGG 5777
 Db 668 rTyrLysGluTTPHISerIysLysGluGluTyrAsnLysGlnAlaLysGluTyrGl 688
 OY 5778 TGAG-----ATATAAGCAAAATATATTCGACATCCCTGTGCAAAAGATGC 5825
 Db 688 nGluTyrGlnLysGlyAsnAsnTyrLysMetTyrSerGlnPheLysSerIleLysProGl 708
 OY 5826 AGAGAGCTCCGCAATATTTAGACAAACAATTAATAATTTGTGAATAATAAGTGC 5885
 Db 708 uValTyrLeuLysLysTyrSerGluLys-----CysSerAsnLeuAsnPh 723
 OY 5886 AGATTGTGAATATAG-----TGT-- 5904
 Db 723 eGluAspGluPheLysGluLysLeuHISerAspTyrLysAsnLysCysThrMetCysPr 743
 OY 5905 -----ATGAAGATGTGTCCACACAGCATAGATGATGATGATGCAAAATATGCCCGC 5960
 Db 743 oGluValLysAspAlaProIleSer---IleIleArgAsnAsnGluIleThrSerGlnGl 762
 OY 5961 ATCATTAAGATGAACCAAAAGAAAT-----GAAGAAAGTGTATATGTCA 6008
 Db 762 uAlaValProGluLysnThrGluIleAlaHISArgThrGluThrProSerIleSerGl 782
 OY 6009 AGTCCACAGAGTCCACACAGTGTACGAGAGGAAACACCGTCAACCGGATACACTGAT 6068
 Db 782 uGlyProLysGlyAsnGluLysGluLysAspAspAspSer-----Le 797
 OY 6069 ATCAAAAGCAGGCGATCGAAAAAAGAAAGCAAGAAACGCGCGCTCAAAACAGCCGAA 6128
 Db 797 uSerLysIleSerValSer-----ProGluAsnSerArgPro-- 809
 OY 6129 AAAAGTGAATTTCTAACACAGAAATGCGAGACAAACAGAACCCGACAGACAGACA 6188
 Db 810 -----G1 810
 OY 6189 ACAAAACAGAAAACGAAATCAACAGCAACAGCAACAGAAATGTGACGTGGC---ACAAT 6245

```

Db      810  urhraspaalalyasprhserasnleuylsleuylslyaspvalasplleserne 830
QY      6246  GATAAAGCCATCTTTCGATAACAGATAGCAGGGGCGAATGAGGGCTGTATACC 6305
Db      830  tProlysalavalileglyserSerProasn-----Aspsnilleasn 844
QY      6306  AAAAAGCTATGACAAATATCCATAATGGGGTGTATGTAGTAACTTAAGAAATGA 6365
Db      845  -----ValThrIuglInglyaspasnll 852
QY      6366  AAATGCATATGTATGCTCTAGAGAAAAAATATATATATATATATATATATATAT 6425
Db      852  eserGlyVal-----AsnSerlyspProleuSeraspaspvalatpProas 867
QY      6426  AAATATTAAGAACTGAATAATAGCGTGCATATATATAAGCGCTTTATTAATATG 6485
Db      867  pLysylsGluLeuGluaspGlnasnserserpsglusergluThrValVal----- 884
QY      6486  AGCAATAGAACTCAATTTTGTGTTAAATATATATATATATATATATATATATAT 6545
Db      885  -----AsnHisIleSerlySerProSerIleasn 895
QY      6546  TGAATATGCAAAATGGAACAATTCAGATGAATTTAAAGAAATATATATATATAT 6605
Db      895  nGlyaspaspSerGlySer-----GlySerAlatThrValse 907
QY      6606  TGATATTAAGAATATGTTTTTGGAACTGATATTTCTAATGATATAAAATATATAT 6665
Db      907  rGluSerSerSerSerAsnThrGlyLeuSerIleaspaspargAsnGlyAspThrPh 927
QY      6666  AACAAATAGTGAACAAACCATTCCTCAATGAATAATATATAAGAAACAGAGATAAA 6725
Db      927  eValalrghrIngsprThrAlaasnThrGluaspAlleatGlysluAsnAlaAspLy 947
QY      6726  AGATGAAGAATACGTAAATATTTTGGAGAGAAATATAAATATATATTTTGGAGAA 6785
Db      947  sAspGluaspGluysGlyAlaaspGluIngr----- 958
QY      6786  GATATATGATTAACCTTATCATCTACAGACCAAAAC-----GAAAAAGAAA 6833
Db      959  -----HisSerThrserGluSerLeuSerSerProGluGluylsme 972
QY      6834  AATTAGATATATTACAGTACATGACATGACCAAACTGACGCTTCCCTTGAAGACT 6893
Db      972  tleuThrspasnGluGlyGlyAsnSerLeuasn-----HisGluGly 986
QY      6894  TGTAAAAAGCCCCCAATTTTGTAGATGGTTCACAGAAATGGCAGAAAGATTTTGA 6953
Db      986  uValylsGluHisThrserAsnSeraspAsnValGlnGlnInserGlyIleValAsn 1006
QY      6954  GAGGAAGGACAGTTCGTTAAAA-----TTGAGAGCGGG 6986
Db      1006  tAsnValGluysGluLeuLysAspThrLeuGluasnProSerSerSerLeuAspGlu 1026
QY      6987  CTGTAAAGCAATATAGTGTATGTAGTATAT----- 7017
Db      1026  yLysAlaHisGluLeuLeuSerGluProasnLeuSerSeraspGlnaspsmetSer 1046
QY      7018  -----GACGTTAAGACACAGAATGTGCAGAGCGGTGTATACATATCAAA 7064
Db      1046  rProGlyProleuaspasnThrSerGluGluThrThrGluArg-----IleSerAsn 1064
QY      7065  TTTTATTAAGCAATGCAAAATCTGATATGAAGAACAAGAAAGTTCAAAAGGAT-- 7122
Db      1064  nGluThrLysValasnGluaspGluaspGluThrLeuThrLysGluThrGluaspP 1084
QY      7123  -----AAAGATGGCAAAAGTAAAGATTAAGCTTATCC 7151
Db      1084  eValleuLysSerHisMetAsnArgLysSeraspGlyGluLeuLysArgLysGlu 1102
QY      7152  TTTTACGAAAGACATTAAGGAAGCAACATGTGCTCATGATATTTAAACATGAAT 7211

```

```

Db      1103  -----AsnSeraspLeuSerThrValAsnaspGluSerGluaspAlaGluAlaLys 1120
QY      7212  AAAAGCAATATATGCGCAATAGCAATGTCTGTATGCAAAAAACCTTCTGCAACTAC 7271
Db      1120  tLys-----GlyAsn-----AspThrSerGluMetSerHisAsnSerSerGlnHis 1136
QY      7272  AAAAACAACAACAATACATATCCGATCCGATTCATATATATATATATATATATAT 7329
Db      1136  eGluSeraspGlnGlnLys-----AsnAspMetLysThrValGlyAspLe 1151
QY      7330  -----TATGTTCTGAGAAATTTAACAAGTGTAGTCTCTGCACTTCAAAA-- 7377
Db      1151  uGlyThrThrHisValGlnAsnGluIleSerValProValThrGlyGluIleAspG 1171
QY      7378  -----AAGGATCTATGATTTCTATCAAAAAAATTAATCTGAACTTAATATAC 7424
Db      1171  sLeuArgGluSerLysGluSerLysIleHis-----LysAlaGluGluIngrGlu 1189
QY      7425  TATGAATGTGTAGAGACAGACATATTTATCTTAAGACAGCAAGAAATATATGCA 7484
Db      1189  rHisThrspIleHisLys-----IleasnProGluaspArgAsnSerAsnThr 1205
QY      7485  TATTACCTTGAAGCAAAATTTATATACCTATTAAGCTTACAAAGGAAAGCAACTAA 7544
Db      1205  rLeuHisLeuLysasp-----IleArgasnGluGluAsnGluArg 1218
QY      7545  TAGTTGACATATATATAT 7563
Db      1218  gHisLeuThrAsnGlnAsn 1224

RESULT 6
US-09-924-154-17
: Sequence 17, Application US/09924154
: Patent No. US20020127241A1
: GENERAL INFORMATION:
: APPLICANT: Natum, David L.
: TITLE OF INVENTION: Anti-plasmodium Compositions and Methods of Use
: FILE REFERENCE: 05213-0465 43170-262105
: CURRENT APPLICATION NUMBER: US/09/924,154
: PRIOR APPLICATION NUMBER: 2001-08-07
: PRIOR FILING DATE: 2000-08-07
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 17
: LENGTH: 1501
: TYPE: PR
: ORGANISM: Mammalian
: US-09-924-154-17

Alignment Scores:
Pred. No.: 8, 01e-31 Length: 1501
Score: 607.00 Matches: 345
Percent Similarity: 34.378 Conservative: 217
Best Local Similarity: 21.108 Mismatches: 548
Query Match: 3.12% Indels: 525
DB: 10 Gaps: 71

US-10-087-013-1 (1-10628) x US-09-924-154-17 (1-1501)
QY      6153  AATGCGAGCAACAACGAAACCGAGACAGCAACAACGAAACGAAACATCAAC 6212
Db      42  AsnValGluThrAsnAsn-----AsnAsnAsnAsnAsnAsnAsnSerAsn 58
QY      6213  AGCAACAACAACAATCTGACGTGGGACAAATGTAAGGCGCATCTTTCGATAAACC 6272
Db      59  SerAsnaspAlaMetSerPheValasnGluValIleArgPheIleGluasnGluLys 78
QY      6273  AGAT----- 6276
Db      78  pAspLysGluaspLysValylsIleIleSerArgProValGluAsnThrLeuHisAr 98

```


QY 4626 TAAATATTACAGATGACGCGCAAAAAAGATTCGATAGACACACATGTGGTAAAT 4685
Db 381 rlysgluluglulhrghlnlysalprolysglu----- 392
QY 4686 GGTACAGACATATCTGGAAGCAATGCAACAGATTTACTGACGAAATTTACTGCTAG 4745
Db 393 -----AsnAlaGluAsnTyrLeu---llelyslleSerGlu 404
QY 4746 TTGTGTATAGAGCTGGAAGTGCCTGTGTGTACAAAGAAATATACATTTGTAGAAA 4805
Db 404 nlysalasap-----AlaLysValSerleuLeuLeu 415
QY 4806 ACAGGCTTACTATGATCCGACAAACATTTGGTGC-----ACAATTTATTGGA 4856
Db 415 nasncysaspAlaGluTyrSerlystYrCysarCyslyshisthrThrlleuVally 435
QY 4857 A-----AATGAC-----GACAAATATACATTTGCGATA 4889
Db 435 sSerValleuAsnGlyAsnspasnthrllelysglulysatgclunhislleaspleuAs 455
QY 4890 AGATTAAGTCGAAA--GATTTAGTAAGAGGCA--AACACAGTGTATTAAAGTGGCA 4943
Db 455 paaspheserlyPheGlyCysasplysasnservalaspthrAsnthrllysalTrrpGl 475
QY 4944 AACCAAGGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5003
Db 475 uclysllyspPro-----TyrLysLeuSerthrllysalpvalCysValPr 490
QY 5004 TTCTGCTGACTAGATATATGTTTCATGATTCATGATTCATGATTCATGATTCATGAT 5063
Db 490 opratagatgnglnlulhrghlnlysalnlysalpalytYr----- 505
QY 5064 TAAAGTGAATGGTGGT-----CGAAAAGATTGATGAGAGTGGCGCGCAAGG 5117
Db 506 -----AsplysasnthrlleuMetllelysglulhislleuAlaIleAlaIleTyrGlu 524
QY 5118 GTACAAATTTGGTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 5177
Db 524 rArglleuLysatgylstYrlysalnlysalpalytYr----- 538
QY 5178 TCCGCAACAATTTCTTTATGAGGTCGCCGCTGTAGTGTATGAAATATGATTTTATGA 5237
Db 539 -----ValCyslysllelleasnthrllysalpalytYr----- 549
QY 5238 TTTAAGATATATTTAGTATGATTAATTTGGAATGAAACAAACAAACGCGAGCA 5297
Db 549 pIleatrgspIlelleGlyglYthAspTyrTrrpAsnsp-----LeuSerAsnAr 566
QY 5298 AAATTTGAAGAAATTTTAAACAAATGGAACATCAGTTGCAAGAGAGATGATGATGAC 5357
Db 566 gLysleuValGlyLyslleasnthrllysalnlysalpalytYr----- 585
QY 5358 TACAGGAATCCCGTACTACTGCGCGCAAAATTTTCTGGAACAAATTAAGATGTGT 5417
Db 566 -----AsplyslleuPheatrgspglutrrpTrrpYsalIlelysllysalpva 601
QY 5418 GTGGAACCAATGATATCGGGTACAAACGTGTAGGATGATGGAATTAAGTGAATAG 5477
Db 601 lTrrpAsnValIleSertrpValpHeLys----- 610
QY 5478 TGCAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5537
Db 611 -----Asplysth 613
QY 5538 TATGGAAGAAATGCGATGAGTACTGGGTATCCTTCGATGCTTTGGCCGATG 5597
Db 613 rValCyslysalnlysalpalytYrleuPheatrgTrrpPheSerGlu 633
QY 5598 GGTGAAGATTTTGAACAAATTAAGAAAGATTTGAGAAATTTGAGGCGGTGTA 5657
Db 633 pGlyaspAspTyrCysGlnAspLysThrLysMetlleGluThrleuLysValGluCysly 653
QY 5658 TCATTTACTGTGTGATTAATGAATTAAGAAAGAAATGATGATGCTGTACACA 5717
Db 653 sgLysProCysGlnAspAspAsnCyslysalYr-----CysAsnSe 668
QY 5718 ATATTAATTTTATGAGTGGGAAACCAAGATGAAACAAATCAATCAATCAATCAAT 5777
Db 668 rTrrpYrsglulrrpIleSerlysllysglululrrpAsnthrlleuVallyGlnTyrGlu 688
QY 5778 TGA-----AATTAAGCAAAATATATTTCCAGCATCTCTGTGGCAAAAGATGC 5825
Db 688 nglulTyrGlnlysglyAsnAsnthrllysalMetlyserGluPheLysSerllelyspProGl 708
QY 5826 AGAGACGCTCCCGCAATTTTGAACAAACAAATTAAGAAATTTGTAAGAAATGAG 5885
Db 708 uVallyrleuLyslystYrSerGluLys-----CysSerAsnthrlleuAsnPh 723
QY 5886 AGATTTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5945
Db 723 eglunspglulhrghln----- 728
QY 5946 TCAAAATATGCCGATCATTAGACGATGAACCAAAAGATGGAAGAAATGATTAAT 6003
Db 729 -----GluGluLeuHlSerAspTrrpYrlysalnlysalYrPhe 741
QY 6004 -TGTCAAGTGCACGAGTCCACACGTCACGAAGGAAACACCGTCACCGGATGC 6062
Db 741 tCys-----ProGluVal--LysaspValPro-----IleSe 751
QY 6063 ACTGATATCAAAAGCAGCGGATGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAG 6122
Db 751 rIlelleatrgAsnAsnGlnGlnThrSerGlnGlnAlaValrrpGluLysThrGlu 771
QY 6123 GCCGAAATGGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6173
Db 771 eAlaHlAsnthrlulhrghlnAsnthrllysalpalytYrleuPheLysAsnthrllysal 791
QY 6174 CCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 6233
Db 791 pLeuLysAsnthrlleuValleuGlnlysalpalytYrleuLysAsnthrlleuLys 811
QY 6234 CGTGGCAATGATGATTAAGGCAATTTCTTTCGATTAACCAAT-----AGCAGGG 6284
Db 811 pLeuGlyGly-----SerArgSerGlnAspGluValThrGlnLuph 825
QY 6285 TGAATGAGGGTGTATTAATCAAAACGATGACAAATATTCATATGCGGTGTATGT 6344
Db 825 eglYvalAsnHlsglylleProlysglylAspGln-----ThrLe 839
QY 6345 AGTAACTTAA-----GAAAATGAATGAGCATGATGATGCGCTCC 6386
Db 839 uglYllysalpAlaIleProAsnthrlleuGluProGluThrGlylleSerThrtrGlu 859
QY 6387 TAGAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6446
Db 859 uglulserArg-----HisGluGlnGlyHisAsnly 869
QY 6447 GCGTGAACATGATTAAGAGGCTTTTATTAATGTCGACAAATGAACCAATTTT 6506
Db 869 sgln-----AlaLeuSerThrSer----- 875
QY 6507 GTGTTAAATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6555
Db 876 -----ValAspGluProGluLeuSerAspThrGlnleuHlsglulns 890
QY 6556 -----AATGAAACATTCAGATTAATTAAGAAATTAATTAATTAATTAATTAAT 6605
Db 890 pThrlysgluAsnAspLysleuProleuLysSerthrThrleuPheProThrGlu 910
QY 6606 TGAATTAAGATGATTTTGGAACTGATATTTCTAATGATTAATAAATTAATTAATGT 6665
Db 910 rglYserSerAsp----- 914
QY 6666 AACAAATAGTGAACACATTTCTCAATGAATAATTAAGAAACGATTAATAAATAA 6725

Db	915	-----ThrGluGluThrProSerIleSerGluGlyProLysGlyAsnGluGlnLysGlyAsp	933
Qy	6726	AGATGAAAGAA-----TTACGTAAATATTTTGGGAGAAAAATTAATAATTTATTTGGGAAG	6782
Db	933	gaspAspAspSerLeuSerLysIle-----	941
Qy	6783	AATGATATATGAGTAACTTATCATCTCTCACAGACGAAACGAAACGAAAAAGAAAAATTTAGGA	6842
Db	942	-----SerAlaIleVal-----	953
Qy	6843	TAATTAACGATACATGACATGACGAAACGACGACGCTTCCCTTGAAGAGCTTGTAAAG	6902
Db	953	pAluLysAspThrSerAsnLeuLeuLysGlyAspValAsp-----IleSerMet	971
Qy	6903	GCCCCAATTTTGGAGATGGTTCACAGAAATGGCGAGAAATTTTGTAAATAGAGAGGA	6962
Db	971	tProLysAlaVal-----IleGlySerProAsnAsp	982
Qy	7023	TAAAGACAAAGAAATGTGCAGAGCGGTGTGACATATCAAAATTTTATTAGAAAGTGGAA	7082
Db	998	-----	998
Qy	7083	AACGTAAATATGAAAGACAAAGAGAAAGTTCCAAAAGATTAAGATGGCAAAAGATATA	7142
Db	999	-----LysProLeuSerAspLysValAlaArgProAspLysAsnHisLysGluValLys	1015
Qy	7143	GGATTTATCTTTCTACTGCAAGAGACATATAGAAAGGCAACATGTCTCATGAATATTATAA	7202
Db	1015	sgLHisIleThrSerAsnSerAspHisValGlnGln-----SerGlyGlyIleValAs	1032
Qy	7203	CATGAATAATA-----AAGAAATATGTGGCAATATGAGATGTCTTGTATGCAAAACCTTC	7259
Db	1032	meAsnValAlaGluSerLysLeu-----LysAsp-----ThrLeuGluAsnProSe	1047
Qy	7260	TTCAACATACCAAAAACACACACACACATCCACATATCCGATGCTAAAT-----	7308
Db	1047	rSerSerLeuAspGluLysAlaHisGluGluLeuSerGluProAsnLeuSerSerAs	1067
Qy	7309	-----GATATG-----CCAGATTCGCTGATTAATGTCTCTGAGAAATTTAACAACTG	7355
Db	1067	pGlnAspMetSerAsnThrProGlyProLeuAspAsnThrSerGlu-----	1083
Qy	7356	TGATGTGTCGTGAACCTTCCAAAAAGGAGTCTATGATTCATACAAAAAAATTTACTGAAC	7415
Db	1084	-----ThrThrGluAlaGlyIleSerAsnAs	1091
Qy	7416	TAAATATACCTATGATATGTGTAGAGAAAGCAGCATATTTATTATCTAAAGAGCGAATA	7475
Db	1091	ngLysThrLysValAsn-----GluAlaGlyGlu	1100
Qy	7476	TAATATGGAATATTAACCTTGAAGAGAAATTTATACCTATTTAGCTATACAAAGAGAAAG	7535
Db	1100	uGluAlaArgThrLeuThrLysGluLysArgLysIleValLeuLysSerHisMetAsnArgG	1120
Qy	7536	AAGTAAATAT-----AGTTGGACATATAATAATCC	7565
Db	1120	uSerAspAspLysGluLeuLysArgLysAsnSerAspLeuSerThrValAsnAspGluSe	1140
Qy	7566	TTTGAGATGCTTAATG-----AAACCTATGACACCTGTAAATATATAGAGAAAGAACCCCTTG	7622
Db	1140	rgLusAspAlaGluAlaLysMetLysGlyAsnAspThrSerGluMetSerHisAsnSerSe	1160
Qy	7623	TGAATAATAGACAGAAATATCTTTAAGATGATTAATGTAATGGAATGTTCACAAAATTC	7682
Db	1160	rgLHisIleIleGluSerAspGlnGlnLysAsnAsp-----MetLysThrValGlyAspLe	1178
Qy	7683	AAAGTTCATAGAGAGAAAAAGAGTATGTCTACTCCAAAGAGAGAA-----CATATGTG	7738

```

Db 1178 uG1yThrHnIsvAlGlnsngIuIeserValProValThnGluIuIleasrGlu 1198
Oy 7740 CTTAAGCAATTTAGATGAATTAATTT-----GAAGACTT----- 7776
Db 1198 slenArGluIserIyGluIserIyIleNIslYsAlaGluIuIuIarGluIserNIsth 1218
Oy 7777 -----AAGATTAATTAATTAATTC-----CTTAATAATGT 7805
Db 1218 rAsrIleHIsLysIleAsnProGluAsrParGAsnSerAsnThrIleNIslYsAsrDl 1238
Oy 7806 TCGT-----CGAATCGAGAAATGAGAAAGACATATATTAATAACT 7850
Db 1238 eArGAsnGluIuAsnGluIuArGhtIsleuThrAsnGlnAsnIle-----AsnIl 1254
Oy 7851 CAACSTCAGAGAACGGGTGGCCAAATGAATCCAAATATGTGATCTAGTAATATAGTTTGC 7910
Db 1254 eSerGluIuArG-----AsrLeuGluNIslYsGluYrheNI 1266
Oy 7911 TGATCTG-----GTCACATAGT----- 7929
Db 1266 sThrMetAsnAsnIleuHIsGluYAsrGluYalSerGluArGserGlnIleAsnHIsSerNI 1286
Oy 7930 -AGAGAAACAGATATGTTCAGGAATGTGGTGTACTACTCCCGTAGAAATTAATTA 7988
Db 1286 sHIsGluYAsnArGlnsParGrlGylGlyAsn----- 1296
Oy 7989 TAAAGTTTTGAAATACATATATATGAAAAATGAGAAATTAATAAAGGTAGAAATTAATA 8048
Db 1297 -----SerGluYAsnValIleuAsnMetArGserAsnAsnAsnAsnRh 1310
Oy 8049 CAACGATGTACAAACGTTTGGTTCTGCTTGCTGGTGGATGCTAATAGAAAAGATATTGGAA 8108
Db 1310 eAsnAsnIle----- 1313
Oy 8109 AGCAATGACGCGAAGACACACAGATGCGAAACTTTT---AGAAAAGAAATGGA 8165
Db 1314 -----ProSerArGluYAsnIleuYrGAsrPlYsIyLeuAsrPlenAs 1327
Oy 8166 TGATTTGAAAGCGATTAACATTAAATACAAAGTAAGTGTGACATTAAGACAGATCCACTGT 8225
Db 1327 PlenYrGlu-----AsnArGAsnAsrSerThrTh 1337
Oy 8226 TGATGATTAATATACCTCAACGGTTTGCATGGATGACTGAAATGGTGAATTAATTGPA 8285
Db 1337 rLyGsluIleuIle-----Ly 1342
Oy 8286 AGCACTGATGGA-----GATTCGAAAAATTTAAAAATCATCTGATCAC- 8331
Db 1342 sLyIleuAlaGluIleAsnIyCysGluAsnGluIuIleSerValIyYrCysAsrHIsme 1362
Oy 8332 -----TGAAACATGTGACAGATCCAGAAATGATTAATGAGAA 8372
Db 1362 tIleHIsGluIuIleProIleuYsThr-----CysThrLyGsluYsThrArGAs 1379
Oy 8373 TAACTGCAACAGTGTAAACAGAGACTCAAGATATAAAAATTTTGTCTTAATAAGAA 8432
Db 1379 nIleuYsCysAlaValSerAsrPlYrCysMetSerYr----- 1391
Oy 8433 ATCTCTATTCGATATACAAATAAATATTC-----AAAGATTTATAGACA 8488
Db 1392 -----PherThrYrAsrPerGluIuYrYrGAsnCysThrLyAsrGluParAs 1409
Oy 8481 ACCAATATATACA 8493
Db 1409 pProSerYrThr 1413

RESULT 8
US-09-924-154-14
: Sequence 14, Application US/09924154
: Patent No. US20020127241A1
: GENERAL INFORMATION:
: APPLICANT: Slim, Kim L.

```

RESULT 8
US-09-924-154-14
; Sequence 14, Application US/09924154
; Patent No. US2002012741A1
; GENERAL INFORMATION:
; APPLICANT: Narum, David L.
; APPLICANT: Sim, Kim L.

; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 ; FILE REFERENCE: 05213-0465 43170-262105
 ; CURRENT APPLICATION NUMBER: US/09/924,154
 ; CURRENT FILING DATE: 2001-08-07
 ; PRIOR APPLICATION NUMBER: US 60/223,525
 ; PRIOR FILING DATE: 2000-08-07
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 14
 ; LENGTH: 1143
 ; TYPE: PRF
 ; ORGANISM: Mammalian
 US-09-924-154-14

Alignment Scores:

	Pred. No.:	Length:	Matches:
Score:	597.00	286	1143
Percent Similarity:	36.11%	Conservative:	195
Best Local Similarity:	21.47%	Mismatches:	432
Query Match:	3.07%	Indels:	419
	10	Gaps:	59

US-10-087-013-1 (1-10628) x US-09-924-154-14 (1-1143)

QY 6006 TCAAGTGCACGAGCCACACGCTGTACGAGGAAACCCGTACACCGGGTATTCAC 6065
 Db 45 SerAspIleSerArgValAsnSerProGluLeuAsnAsnHisLysThrAsnIleTyr 64
 QY 6066 GATATCAAAAGGACGCGATCGAAAAAGAGCAAAACGCGCGCCGTACAAAAACGCC 6125
 Db 65 AspSerAspTyrGluAspAlaAsnAsnLysLeuIleAsnSerPheValGluAsnLysSer 84
 QY 6126 GAAAAAGTGGAAATCT-----AACAAAGAAATGCGAGCAAAACAGCAAC 6173
 Db 85 ValLysLysArgSerLeuSerPheIleAsnAsnLysSerLysSerTyrAspIleIle 104
 QY 6174 CCGAGAGC-----AGC 6185
 Db 105 ProProSerTyrSerTyrArgAsnAspLysPheAsnSerLeuSerGluAsnGluAspAsn 124
 QY 6186 ACAACAAACAGCAAAACGACATCAACAGCAACAAACAGAAATCTGACGTGGCAGCAAT 6245
 Db 125 SerGluAsnThrAsnSerAsnAsnPheIleAsnThrSerGluIleSerIleGlyLysAsp 144
 QY 6246 GGTAAAGCCATCTCTTGTGAATTAACACGATGACAGGGGTGATGAGGTTGATCC 6305
 Db 145 AsnLysGlnTyrThrPhe-----IleGln-LysArgThrHisLeuPheAlaCysGlyI 162
 QY 6306 AAAAACGATGACAAATATCTAAATGGGGTGTATGTAGTAACTTAAAGAAATGA 6365
 Db 162 eLysArg-----LysSerIleLysTrpIleCys-----ArgGluAsnSe 175
 QY 6366 AAATGGCATA-----TGATAGCTCTAGAGAGAAAAATTTATGTTAAATAATATACA 6419
 Db 175 rGluLysIleIleThrValCysValProAspArgLysIleGlnLeuLysIleAlaAsn 195
 QY 6420 ATATTAAATTAAGAACTGAAATTAAGCGTGAACATGATATTAAGAGGCTTTTATTA 6479
 Db 195 uAsnSerArgLenglunThrMetGluLys-----PheLysGluIlePheLeuI 211
 QY 6480 ATGCGACGCAATAGAACTCAATTTTGTGGTTAAATATATATGAAATCTGCAGC 6539
 Db 211 eSerValAsnThrGluAlaLysLeuLysLeuTyrAsnLysAsnGluGlyLysAspProSerI 231
 QY 6540 AGAA--AAAGAAATGCAAAATGCAACAAATCCAGATGAAATTTAAAGAAATATATTA 6596
 Db 231 ePheCysAsnGluLeuArgAsn----- 238
 QY 6597 TACATATGCGATATTAAGATATGTTTTTGGAGCTGATATTTCTATGATTAATAAAT 6656
 Db 239 -SerPheSerAspPheArgAsnSerPheIleGlyAspAspMetAspPheGlyGlyAsn 258
 QY 6657 TATAAGTGTAAACAATAGGTAAACAACATTCATGAAATTAATTAAGAAAAACAGCA 6716

Db 258 rAspArgValLysGlyTyrIle-----AsnLysLysPheSerAs 271
 QY 6717 TAAAAAAAAGATGAAGATTTACGTAAATA-----TTTGGAGAA 6758
 Db 271 PLYTyrTrpLysGluLysAsnValGluLysLeuAsnAsnIleLysLysGluTrpTrpGluL 291
 QY 6759 AAATAAAAATTTATTTGGGAAGAAATGATATGATGATTAATCATCATCAACAGACA 6818
 Db 291 sAsnLysAlaAsnLeuTrpAsnHisMetIleVal----- 302
 QY 6819 AAACGAAAAAGAAAAATTAAGATTAATTTACAGTACAAATGACAACTGACGCC 6878
 Db 303 -AsnHisLysGlyAsnIle-----SerLysGluCysAlaIleIlePr 316
 QY 6879 TTCCCTTGAAAGATTTGTAAGAAAGCCCAATTTTGAATGCTTACACAAATGGGACGA 6938
 Db 316 oAlaGluGlu-----ProGlnIleAsnLeuTrpIleLysGluTrpAsnG 331
 QY 6939 AGAATTTTGTATTAAGAGGAAGAACAGTGTAAAAATGAGCGCGGCTGT----- 6990
 Db 331 uAsnPheLeuMetGluLysLysArgLeuPheLeuAsnIleLysAspLysCysValGluAs 351
 QY 6991 -AAGAAATATGAG--TGTAATGCTAGTATATGACGTAAACACAAAGATGTCAGAGCC 7046
 Db 351 nLysLysTyrGluAlaLysPheGlyGly-----CysArgLeuPr 364
 QY 7047 GTGTGATACATCAAAATTTTATTAAGAGTGAAGAAATGATATGAAGAAACAAAGAGA 7106
 Db 364 oCysSerSerTyrThrSerPheMetLysLysSerLysThrGluMetGluValLeuThrAs 384
 QY 7107 AAAGTCAAAAAGATTAAGATGCAAAAAGTAAAGATTAATCTTCTACGAAAGAGA 7166
 Db 384 nLeuTyrLysLysLysAsnSerLys----- 392
 QY 7167 CATAGAGAACGAAACATGCTCATGAATTTAAACATGAATTAAGAAATATGTCG 7226
 Db 393 -ValAspLys-----AsnAsnPheLeuAsnAspLeuPheLysLys----- 405
 QY 7227 CAATTAAGATTTGCTTGTATGCAAAAACCTTCTCAACATCAACAAACAAACACA 7286
 Db 405 ----- 405
 QY 7287 ATCACAATATCCGATGATATGATATGCGAGATCGTGAATTTATGTTCTGAGAAAT 7346
 Db 406 -----AsnAsnLysAsnAsnAspLeuAspAspPhePheLysAsnGluLysGluTyrAs 422
 QY 7347 TAAACAGTGTGAGTCTCGAAGCTTCAAAAAAGGATCTATGATTCATAAAAAAAT 7406
 Db 422 pAspLeuCysAspCysArg-----TyrThrAlaThrI 433
 QY 7407 TACTGAACCTAAAAATACATGATTAATTTGTAGAGAAAGACATTTATTTATCTTAAGA 7466
 Db 433 eIleLysSer-----PheLeuAsnGlyPr 441
 QY 7467 AGCAGAAAAATTAATGATATTAACCTGAAGAAAAATTTATACCTATTGAGTGTACAAA 7526
 Db 441 oAlaLysAsnAspValAspIleAla----- 449
 QY 7527 GAAAAAGAAATTAATTAATGCTGATTAATTAATTCCTTCGATCCTAAGAAACCTTA 7586
 Db 450 -----SerGlnIleAsnValAsnAspLeu----- 457
 QY 7587 TGCACCTGATTAATTAATTAAGAGAAACCCCTGTGTAATAATAGAGAAATTCGTT 7646
 Db 458 -----ArgGlyPheGlyCysAsnTyrLysSerAsnAsn----- 468
 QY 7647 TAAAGTATATGATGAGAAATGTTACAAAATTCAAAGTCTATCAGAGAGAAAAAAG 7706
 Db 469 -----GluLysSerTrpAsnCys-----ThrGlyThrPheThrAsnLysPheProG 484
 QY 7707 AGTATGCTACCTCCAGAAAGAAACAAATATGCTTAAGGAATTTA----- 7752

Db 484 yThrCysGluProIleArgGlnIleuCysLeuGlyArgThrTyIleuLeuHisar 504
QY 7753 -----GATGAATTAATAATTAAGACACTAAGAGATTAATTAATCTCTAAAT 7802
Db 504 gGlyHisGluGlnAspTyrIleuGlnIleuLeuGlyAlaSerIleTyIleuLacGlnle 524
QY 7803 GGTTCCTCGAATGCGACGAATGAGAAATAGCATTAATAAACTCACTCAAGAA 7862
Db 524 uIeuLysTyrIleuGlnIleuLysAsp----- 533
QY 7863 CGGTCGCGAATGATCCATATGATGATCTAATAATAGTTGCGTGCATGCGTGA 7922
Db 534 -----GluAsnAlaLeuCysSerIleIleGlnAsnSerTyrAlaAspLeuAlaas 550
QY 7923 CATAGTAGAGAAAGATATGTTAGCAATGCTGGTACTTACTCCCTAGAAATAA 7982
Db 550 pIleIleuGlnIleuSerAspIleIleLys---AspTyrTyrGlyLysLysMetGluGln 559
QY 7983 ATTATATAGGTTTGGATACATATATGGAATAAGCAAAATAAATAAAGGTGAAA 8042
Db 569 uIeuAsnLysVal-----AsnLysAspLysLysArgas 580
QY 8043 TAATATCAACAGATGTCAAAAGCTTTCCTGCTGCTGGTGGAGATGATGAAGATAT 8102
Db 580 nGlu---GluSerLeuLysIlePheArgGlnLysTyrIlePheGlnLysGlnLysnVa 599
QY 8103 TTGGAAGAAGATG-----ACGTGCAAGACACACAGAGATGC 8138
Db 599 lTrpLysValMetSerAlaValIleuLysAsnLysGlnLysThrCysLys-----AspTyr 616
QY 8139 AAACTTTTGAAGAAAGAAATGATGATTTGAACGATTAACATTAATACAGATTA 8198
Db 616 rAspLysPheGlnLys----- 621
QY 8199 GTGTGACATAGAGAGATCCACCTGTGATGATTAATATCAACGCTTTCAGATGAT 8258
Db 622 -----IleProGlnPheLeuArgTyrPhe 639
QY 8259 GACTGAATGCTGCAATTAATTAATTAAGACAGATGAGAAATGCAAAATTTTAAAA 8318
Db 629 eLysGlnTyrGlnLysAspPheCysGlnLysArgLysIleuTyIleuSerPheGln 649
QY 8319 ATGATGATGATCACTGTAATAACATCTGACAGATGCAAGATGATTAATTAATTAAGT 8378
Db 649 rPheLysValGlnCysLys-----LysLysAspCysAspGlnLysn----- 662
QY 8379 TGAACAGTGTAAACAGATGTCGAATTAATTAATTAATTTGTTCTTAATTAAGAAATCTCT 8438
Db 663 -----ThrCysLysAsnLysCysSerGlnTyrLysIleAspLeuLysLysSerGln 681
QY 8439 ATTGATTAATCAATCAATTAATTAATTAAGAAATTTGATGAACAACATATATTAACAAAT 8498
Db 681 uTyrGlnLysGlnLysAspLysTyrThrLysAspLysAsnLysLysMetLysAspLysn 701
QY 8499 CTGTACTTAATGATCATGTTCAAAATTTTGTACAAAGTTGAAAATTTTAAAGTGAATG 8558
Db 701 e-----AspGlnValLysAsn-----LysGlnVal 709
QY 8559 TTCTGTGAGAGCTTTTCTGAATATCTTCATGAACAAGTAAAG-----TGTGTAATTAATA 8615
Db 709 aAsnVal-----TyrLeuLysGlnLysSerLysGlnLysLysAspValas 724
QY 8616 ATTGAATGAAGATGATGTTCTTCATATAGCAATATGCTTTCGAGAGAAACACCAAA 8675
Db 724 nPheAspAspLys-----IlePheAsnGlnLysSerProas 735
QY 8676 AACTTAATGAAGATGCTGAGTGTACACTACTTGAAGATCCATGGAATTAATGCTC 8735
Db 735 nGlnTyrGlnLysAspMetCysLysLysCysAspGlnLysLysTyr---LeuAsnGlnLysLys 754
QY 8736 TACCGATTAACAACAAGATGATGATGAAGCAATTAACAACATTTT-----ACCTTCTGCTC 8789
Db 754 sTyrProLysThrLysHisAspIleTyrAspIleAspThrPheSerAspThrPheGlnLys 774

QY 8790 GAAGATGATTAATGAATAATCTGTAATTAATGAAACGATACCTTGTCTTAATAGTTC 8849
Db 774 pGlyThrProIleSerIleAsnLysn-----IleAsnGln 787
QY 8850 AGATGATTAACAAGGCTATGATTCCTCCAGAGAGACATTAATGACAGACCTAT 8909
Db 787 nGlnSerGlyLysAspThr-----Se 794
QY 8910 CACTGCATTAATTAAGAAAGGTGATAGAAATTTTAAAAAACTTCTTACTTC 8969
Db 794 rAsnThrGlnAsnSerGlnThrSerAspSerProValSerHisGlnProGlnSerAspAl 814
QY 8970 TGCTTTCAGCAAGACAAATTTAGCTCAAAATTAATTAATGCAAGAGAGTGTGCTT 9029
Db 814 rAlaIleAsnValGlnLysLeuSerGlnLysAspGlnLysSerSerGlnThrArgGlyIle 834
QY 9030 TGAAGCA-----AT 9038
Db 834 uAspIleAsnAspProSerValThrAsnAsnValAsnGlnValHisAspAlaSerAsnTh 854
QY 9039 GAATATATGTTATGACAGATTAATCCGATTAATTAATAAGCA-----ACGTATATGAT 9089
Db 854 rGlnGlnSerValSerAsnThrSerAspIleThrAsnGlnHisSerGlnLysSerLeuAs 874
QY 9090 GGAACACTTCATTAATCTGAAAAATTAATAATTAATTTGAACAATGACAAACGCA 9149
Db 874 nArgThrThrAsnAlaGlnAspIle---LysIleGlnArgSerGlnLysGlnLysSerAs 893
QY 9150 AAATCGTAAACATGCTGGGAAAAATTAATAGACGTCAATATGCGACGCTTGTATGATGG 9209
Db 893 rAsnGln-----GlnLysSerSer-----HisSerSerAspAsn 905
QY 9210 ATTAATAATGCTACTTCAAACAGATTAACATTAAGTAA----- 9246
Db 905 rGlnSerLeuThrIleGlnGlnValProSerGlnAspAsnThrGlnAsnThrTyIleAsp 925
QY 9247 -----GGATGCTCAATTAACAAGGATGAGAAAC 9278
Db 925 rGlnAsnProHisArgAspThrProAsnAlaLeuAlaSerLeuProSerAspLysIle 945
QY 9279 TAATGATTTCTTCCTGCTGTAATTAATGAGGCAAGCAAGCATGTAAGAAAGAAACA 9338
Db 945 eAsnGln-----IleGlnGlyPheAspSerSerArgAspSerGlnLysnGln 960
QY 9339 TGTAAAGTATTCATTAATAAACAATAATGCTCCTCAACAGAGATTAATTTGAAGGTC 9398
Db 960 yArgGlyAspThr-----ThrSerAsnThrHisAspValArgArgTh 974
QY 9399 AGAATTAATTAAGACAACCTGGATGTCGAATGATTAATTAAGAAATTAATTAAGCTGAATAT 9458
Db 974 rAsnIleValSerGln-----ArgArgValAsnSerHisAs 986
QY 9459 ATTGATTAATAATTAATCAATGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 9507
Db 986 rPheIleArgAsnGlnMetAlaAsnAsnAsnAlaHisHisGlnTyrIleThrGlnIleGln 1006
QY 9508 -----AAAGTCAATCTTCAGTAAATTAATGACAAAT----- 9537
Db 1006 uAsnAsnGlnIleIleArgGlnGlnGlnGlnLysSerAlaGlnLysSerValAsnTyrLysAs 1026
QY 9538 -----AAACATCTGAAGAAATGCTGATCATA 9566
Db 1026 rAsnProLysArgSerAsnPheSerSerGlnLysAsnAsnHisLysLysLysIleGlnGln 1046
QY 9567 TATTAATTAACAAGATCTCAATGCGCTTGGAGTTAAATGATTAATTAATTAATTAATTAAT 9626
Db 1046 r---AsnSerArgAspThrLysArgValArgGln-----GlnIleLys 1060
QY 9627 AGGACAAAAAATTAATGAA-----AATATGATGATTC 9657
Db 1060 sLeuSerLysGlnAsnLysCysAsnAsnGlnTyr 1071

D	55	glturtyrsgluasnasnvalspillellephegltyrthleuansntrglutyr-----	72
O	7513	ATTGAGCTCTCAAGAGAAAGAAAGTAAATAAGTTCGACATATATATATATTCCTTGCAAT	7572
D	73	Asnasphecytylsglysntrproglutleuvalseralaialalysntrysn-----	89
O	7573	CCCTAGAAACCTTATGACCTGATATAATATATAGGAAGAAACCTGTGTCAAAATGACA	7633
D	90	leuylsyalproasnlaalysserproargiletyrtylserlys-----glutisglu	107
O	7633	GAAGAAATCGTTTT-----AAGTACATATATGACATGGAATGAT	7671
D	108	gluserservalphegltcystrthlyrlylleerlystvalylslystrpsnocy	127
O	7672	TACAAAATTCACAAAGTCTATCAGAGACAAAAAAGACATATGTCACCTCCAAAGAGAA	7731
D	128	tyrserasnshly---valthrlyserproglutylvaltlysylproproarglgin	146
O	7732	CATATGTGCTTAAGAAATTTAGATGAATTAATA-----ATTGAAGACTTAAGAT	7782
D	147	gluleucylseugltyrlllepheleuileargpolsynsnlglutluleuylasp	166
O	7783	AGTAATATCTCTAAAAATGTTCTGTCGACATCTCACAGAAATGACAAAGAAATACATATA	7842
D	167	His-----leasnlysalalasntrglutlualamethislenlys	180
O	7843	AAAACTTCACCTCAGAGAACGGGTGGCAATGATACCAATATATGATATCTATGCAAAAT	7902
D	181	glutyslyrgluasnalaiglycy---Aspyslilecysasnalaileuugly	197
O	7903	AGTTTCGCTGATCTGGTGACATATGTTAGAAGAACAGATATGTTCAGATTTGGTGGTTAC	7962
D	198	serlyrtrialaspillegllyaspillevalarglyleuasnvaltrparg-----	213
O	7963	TTACCTCCCGTGAATAAT-----AAATTATATAGGTTTGGATATACATATACGA	8013
D	214	-----AspilleasnthrAsnlyslenuserglutylsphenelulysllephebet	229
O	8014	AAATGAGAAATTAATAAAGTAGACATAAATATACACAGATGTCAACGTTTGGCTTC	8073
D	230	glyglglyasnserarglytlysgluasnaspasnnglu-----Argasn	245
O	8074	GCTTGCGGATGCTATATGAAAAAGATTTTGGAAAGCAATG-----	8115
D	246	lystrptprglutylsnglnarghshleuiletrpserSermetvalylshsileprolys	265
O	8116	-----ACGTCAAAAGCACCAAGAAAGATGCAAAACTTTTACAAAAGACAGATGATGAGA	8169
D	266	glylstrthrcyslys-----ArgHisasnasn	274
O	8170	TTTGAAGCCATACATTAATATACAAAGATTAAGTGTGCACATTAAGACAGATCCACTGTGAT	8229
D	275	pheglutyls-----	277
O	8230	GATTATATACCTCAACGGTTTCGATGATGACGTGATGCTGAATATTAATGTGAAGCA	8289
D	278	-----lleproglinpheleuargtrpleuylsglutrpglyasnpgulnphcysgluuglu	295
O	8290	CTGATGCAAGCAATTTGAAAAATTTAAAAATATATGTGATCACTCTAAAAACATCTGCAGA	8349
D	296	Metelylthrclutllyllysglnleuuglyllsilecysgluasnlyasnocyserglutlys	315
O	8350	TGCAAGATGATTATGATGAATAATAGTGTGACACGTGTAAAAAGACAGATGTCAAGAAATAT	8409
D	316	-----lysctysasnalaicysserseityr	324
O	8410	AAAAATTTGTTCTTAATAGGAATCTCTATTCGATATATACATCAATAATATCAAAAGAA	8469
D	325	glulustrpillelysgluarglyasnolglutyrasnleuuglnserlysnpheser	344
O	8470	TTGTATGAACAACCAATATATACAAAAATTCCTCTATATGATGATATGTTCCAAATTTTGTGA	8529
D	345	-----Aspysllystleuasnlyslsasnshsnleuutyrasn-----	356

```

OY 8530 CAAAGCTGAAACCTTTAAAGTGAAGTCTGTTGAGAGCTTTTCTGAATATCTCAT 8589
DB 357 -----LysPheGluAspSerLysAlaTyrLeuArg 366
OY 8590 GAAACAGTAG--TGTTGAATTATAATTAAGAAAGATGCTTCCATATA 8646
DB 367 SerGluSerLysGlnCysSerAsnIleGluPheAsnAsp----- 379
OY 8647 CGAACATATGCTTTCGAGAACACCAAAAGTATTAAGAACCTTGCAGT--TGTACA 8703
DB 380 GluThrPheThrPhe-----ProAsnLysTyrLysGluAlaLeuGlySerMetValCysGlu 396
OY 8704 CTACTCTTAAGAAAT-----CCATTGGATTAAT--TGCTCCACCGATCA 8745
DB 397 AsnProSerSerSerLysAlaLeuLysProIleLysThrAsnValPheProIleGluGlu 416
OY 8746 AACAAAGATGATGATGAAGATTAACAACCTTTCCTTCTGCTCCGAGAAATGATTATGAT 8805
DB 417 SerLysLys-----SerGluLeuSerSerLeuThrAspLysSerLysAsn----- 431
OY 8806 AATTAATCTGATTAATGGAACCATCTGTTCTTAATAGTTCAGATGATAACAAAGGT 8865
DB 432 -----ThrProAsnSerSerGlyGly----- 439
OY 8866 GTATTGATTCCTCCAGAAAGAACATTTATGTACAAAGCTATACGATATTAATAT 8925
DB 440 -----AsnTyr 441
OY 8926 AGAAAGATGATTAAGAAATTTTAAAGAA-----NAACTT 8961
DB 442 -----GlyAspArgGlnIleSerLysArgAspAspValHisHisAspGlyProLysGlu 459
OY 8962 CTACTCTGCTTTCAGTCAGACAAATTTAGTCAAAATATTAATCGAAGAACAG 9021
DB 460 ValLysSerGlyLysGluValProLysIleAspAlaIleValLysThrGluAsnGlu 479
OY 9022 TTGTGC-----TTGAGCGAAGAAATAGTATTCAGATTAATTCAGAT 9066
DB 480 PheThrSerAsnArgAsnAspIleGluGlyLysGluLysSerLysGlyAspHisSerSer 499
OY 9067 ATATATTAAGAACTGATATGATGACACTTCATTATCGAAAAATTAATAAAATATTT 9126
DB 500 ProValHisSerLysAspIle-----LysAsnGlu 509
OY 9127 GAAACATCAATGAAGCAACCGAAATCGTAAACATGCTGGGAAATATAGACCTCAG 9186
DB 510 GluProGlnArgValValSerGluAsn----- 518
OY 9187 ATATGCGACGCTATGTATTGATTAATAAATTTCTACTCAAAAGTAAACATTAGATGAA 9246
DB 518 ----- 518
OY 9247 GAGTGTGTCATTACCAAGATGAAGAACTAATCACTTTCTTGTTGTTAATTGAA 9306
DB 519 -----LeuProLysIleGlu----- 523
OY 9307 TGGGCAACCAAGCATTAAGGAAAGAAACATGTAAGTATTCATTAACAAATATGT 9366
DB 524 -----GluLysMetGluSerSerAspSerIleProIle----- 534
OY 9367 CCTCGTTCAAGAGATATTTTGAAGCGTCGAATTTTAAGACACCTGGATGTCAG 9426
DB 535 -----ThriHisIleGluAlaGlu----- 540
OY 9427 AATGATATTAGAAATATATTAGCTTGAAATATATTGATAAAATATCAATGAAATCTA 9486
DB 540 ----- 540
OY 9487 AATATATAATATAAGCAATTAAGATCAATCTTCAGTAATATAGACATTAACCATCT 9546
DB 541 -----LysGlyLysSerSerAsnSerSerAspAsnAspProAla 553

```

```

OY 9547 -----GAGAA 9552
DB 554 ValValSerGlyArgGluSerLysAspValAsnLeuHisThrSerGluArgIleLysGlu 573
OY 9553 AATGTCAGGCAATATTAATAATCAAAAGATTCATGGCGCTTGGAGTTA----- 9603
DB 574 AsnGluGluGlyValIleLysThrAspAspSerSerLysSerIleGluIleSerLysIle 593
OY 9604 -----AATGATATAATGAATGATGTTACAGCAACAAATAATAT 9642
DB 594 ProSerAspGlnAsnAsnHisSerAspLeuSerGlnAsnAlaAsnLysAspSerAsnGln 613
OY 9643 GAAATATATGA-----TTCAAAGAGTACTAATAAAATATATATCT 9684
DB 614 GlyAsnLysGluThrIleAsnProProSerThrGluLysAsnLeuLysGluIleHis 632
OY 9685 GGTATTATTTTGTTCAGATGAAACACCAAAATCATGTAATGATGAAATATATAA 9744
DB 633 -----TyrLysThrSerAspSerAspAspHisGlySer 643
OY 9745 GAGAGAGCAACAGTTCGCTCTTAAGCACCTATTTCTTTACACCCATGTAGATTC 9804
DB 644 LysIleLysSerGluIleGluProLysGluLeuThrGluLysSerProLeuThrAspLys 663
OY 9805 TTCTATCAAGACCTTTAATTCACACATCGAGTACACAAATATATGAT 9864
DB 664 -----LysThrGluSerAlaAlaIleGlyAspLysAsnHisGlu 676
OY 9865 ATATGAAAGTACT----- 9879
DB 677 SerValLysSerLysAspIlePheGlnSerGluIleHisAsnSerAspAsnArgAspArg 696
OY 9880 ---ATCTGCTGCTTATTTGTA-----TCGGCGTTCAGTTTGATAGCGCTT 9921
DB 697 IleValSerGluSerValValGlnAspSerSerGlySerSerMetSerThrGluSerIle 716
OY 9922 CATTTATGAGAAATAATCAATGCTGTGAGCTTGTGGATATGAT 9975
DB 717 ArgThrAspAsnLysAspPheLysThrSerGluAspIleAlaProSerIleAsn 734

RESULT 10
US-09-924-154-16
; Sequence 16, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
; APPLICANT: Natum, David L.
; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924/154
; PRIOR APPLICATION NUMBER: US 60/223,525
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
; LENGTH: 972
; TYPE: PRF
; ORGANISM: Mammalian
US-09-924-154-16

Alignment Scores:
Pred. No.: 1,81e-16 Length: 972
Score: 382.50 Matches: 253
Percent Similarity: 31.86% Conservative: 157
Best Local Similarity: 19.66% Mismatches: 450
Query Match: 1.97% Indels: 49
DB: 10 Gaps: 49

US-10-087-013-1 (1-10628) x US-09-924-154-16 (1-972)
OY 3151 TTAAGTGTGATGCGCAATGGCAGAAATGCTACTGCAAGTACGAGAAAGAGATATGAT 3210
DB: 3151 TTAAGTGTGATGCGCAATGGCAGAAATGCTACTGCAAGTACGAGAAAGAGATATGAT 3210

```

[illegible]

[illegible]

230 MetIHTGIUTTPALAGLUTPTPTCYCSLYSALAGLInserGInGLUTTYRASPLYSLeuLYS 249

9 Gaps: 60

PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12996
 LENGTH: 6281
 TYPE: PRF
 ORGANISM: Staphylococcus aureus
 US-09-815-242-12996

Alignment Scores:

Pred. No.:	2,77e-11	Length:	6281
Score:	304.50	Matches:	722
Percent Similarity:	30.47%	Conservative:	529
Best Local Similarity:	17.58%	Mismatches:	1574
Query Match:	1.57%	Indels:	1282
	10	Gaps:	173

US-10-087-013-1 (1-10628) x US-09-815-242-12996 (1-6281)

```

31 AAATGGGAGATGACATCATCTAGAGAGATGCTAAAGCCCTATTATAAAGAA 90
    ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1974 GlnValasnsrLalysSerLalaleuAnslYAspLulysLeuAlaAlaLysGln 1993
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 91 AGTCACAAAAGT-----GCAAGAAATGTTTGGACGTTATGCCAA 132
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1994 ThrAlalysSerAspLleGlYArgLeuThrAspLeuAnslAglInAlaAla 2013
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 133 AAATATAAGATCATCATCAAAATATGCAAAAGACATCGATTCGTTGAAGCGATTG 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2014 AsnAlaGlYValAspGlnAlaProAsnLeuAlaValAlaThrAlaAlaLysAla 2033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 193 AGCAAGACAGAAATTCGTGTGTCTCTACGCCAGTAATATAGCATATATTATTAT 252
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2034 ThrSerLeuAnslThrAlaMetGly----- 2041
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 253 CCATATCATCATGTAATTTAGTCAT-----AAGACATCATCTAATTTACGCTAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2042 -----AsnLeuLysHISAlaLeuAlaGluLysAspAnslThrLys----- 2054
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 301 GATGATGTGAATTTGACACATCCTGCCATGATGAGAAACAAACCGATTGTGAAGAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2055 ----- 2063
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 361 GAAGAAATCGAATGTGGAATAAATACTAATTAATAAAGAAATAATGATGATAGCC 420
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2064 GlnPro-----LysGlnGlnAlaLysAsp 2071
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 421 TGTGCCGCCACCTGAGAGACATATGTGTGATPAAAAAC----- 459
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2072 ThrAlaValThrGlnAlaGlnAlaLleThrAsnAlaAnslYSerAnslAlaAnslYThr 2091
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 -----TTGSAAGCTCTAATGATATTAATACCAAAATTCATGATGTTATGGCA 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2092 GlnValAlaGlnAlaLeuAnslGlnLeuAnslGlnAlaLysAnslPLeuAnslYAsp 2110
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 511 AATGCTACTGATACAGCAAAATAGCAAGATGATCAATTTGATATATCATCATCAATAA 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2111 AsnLysValAlaGlnAlaLysGlnSerAlaLysArgAlaLeuAlaSer-----TyrSer 2128
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 571 GGAATTCACAGCGCTGTACTGCTTGGCAGAACTTTGCGAGATATAGCTGATATTGTA 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2129 AsnLeuAnslAnslAglInsThrAlaLleThrSerGlnLleAspAnslAlaThrThrVal 2148
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 631 AGAGAAATATATATGTTTAAACCAATGTCCATGACAAAGTACAAAGCGGTCCCGAG 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2149 AlagLysAlaThrAlaAlaGln-----AsnThrAlaAnslYLeuAnslThrAlaMetGlyGln 2167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 691 GTTTTCAAGAAATATACATGATGATGAGAGATGAAATGATGATGATGATGATGATGAT 744
    ::| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2168 LeuGlnAnslGlnLysLeuAnsl-----GlnAnslThrValLysGlnGlnValAnslPheThr 2185
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

```

QY 745 -----CTGATGATCTGCAAAATTATTAT----- 768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2186 AspAlaAspGlnGlnLysLysAspAlaLysThrAsnAlaValThrAsnAlaGlnGlyLe 2205
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 769 -----AAATTAGAGAAAGCATGCGGAGATGATGATGATGATGATGATGATGATGAT 807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2206 LeuAspLysAlaHISGlnAnslMetThrLysAlaGlnValGlnAlaAlaLeuAnslGln 2225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 808 GTATGG-----GAGCTAATACATGTGATGATGATGATGATGATGATGATGATGATGATGAT 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2226 ValThrThrAlaLysAnslAlaLeuAnslGlnYAspAlaAnslValArgGlnAlaLys-----Ser 2244
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 859 CAATCAGAAAGTAATACACCATTTATTTCAATTCCTAATGCGCGCATTAACAA----- 912
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2245 AspAlaLysAlaAnslAnslGlnThrLeuThrHISLeuAnslAnslAglInLysGlnAspLeu 2264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 913 -----GGAAGGTTCCACCAAT----- 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2265 ThrSerGlnLleGlnGlnAlaThrThrValAsnGlyValAsnGlyValLysThrLysAla 2284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 931 -----TTAGATTATGTCCTCAATATATTACGTTGTCGACGAATGGGAGAAAGTTT 984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2285 GlnAspLeuAnslPLeuAlaMetGlnArgLeuGln----- 2295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 985 TGCAGAAAGAAATATTAATTAATGAAAAAGCTACCTGCTGCTAATGACAAAGAA 1044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2296 -----SerAlaLleAlaAnslYAsp 2302
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1045 CGCTTATATGTGATCATATATGACATGATGTAACGACATTTTGGAAAAAGATAT 1104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2303 GlnThrLysAlaSerGlnAnslYrLleAspAlaAspProthr-----LysLys----- 2318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1105 TTGATTTGGATATAATAGTACTGACTGCTGCACTAATGCAAAAGTTTGAAGTTTG 1164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2319 ThrAlaPheAspAnslAlaLleThrGlnAla-----GlnSerTyr 2331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1165 TTAGCGAATCAACAAGACATTTAAAAACAAAGAAATAATGCAAAAGAAATACAA 1224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2332 Leu-----AsnLysAspHISGlnAlaAnslYAspLysGlnAlaValGlnGlnAlaLleGln 2350
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1225 TCATATTATTCGAACGATCAAAATTTGCAATTAATTAATGTAATTTTAAACAA 1284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2351 SerValThrSerThrGlnAnslAlaLeuAnslYAspAlaAnslLeuGln----- 2366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1285 TTTTATGAAAAAGTATAGGAAGCAATATGCAACTATGACATTTTAAATTTA 1341
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2367 ---ArgAlaLysThrGlnAlaLleGlnAlaLleAspAnslLeuThrHISLeuAnslThrPro 2385
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1342 -----CTAATGAAAGAAAGTATGTAAGAGAGATTACCA 1377
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2386 GlnLysThrAlaLeuLysGlnGlnValAlaAnslAlaGlnArgValSerGlyValThrAsp 2405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1378 GGAGAAAGATATATCTTTTACTTACACAGCGCATGATGACAAAGGATATTTATCGTTCA 1437
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2406 LeuLysAnslSerAlaThrSerLeuAnslAnslA----- 2416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1438 GAATATTGCCAAGTGTCCCGACTGCGGGGTCAAAATGTGATGATTAATAATACACACAC 1497
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2416 ----- 2416
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1498 AATACATATATGATGCTGTAACGCTTAATTAATGACACTTAATTAACCTCATGGGGTGTG 1557
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2417 ---MetAspGlnLeuLysGlnAlaLleAlaAspHISAsp----- 2428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1558 AAGCTCTAATATCACTGCTTATAGTGAATACACAAAGGTGATATTACCAAAAA 1617
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2429 -----ThrLleValAlaSerGly----- 2434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1618 TTAGAAAAATTTTGTACACACTCAACTAATTAACAAAGATTAATAATTAATTAATTAATTAAT 1677
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2435 -----AsnTyrThrAsnAlaSer-----ProAspLysGlnGlnAlaLysThrAsp 2449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1678 TGCTATTATTAAGATGCAAAATATTAATGATGTAATGATGCAACAAATTAATGATGATGAT 1737
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```

Db 2450 AlaTyrAsnAlaAlaLysAsn-----IleVal 2458
OY 1738 AATGATTAATCCCTAAGTATATATATCATTAATTTTGAATATATGGCTTACATATTTA 1797
Db 2459 AsnGlySerProAsnValIleThr-----AsnAlaAlaAspValThrAlaAla----- 2474
OY 1798 TTAAAGGATCTATTAACTGGAATGACAACCTTAAACTGTGTAATAATATACAAACG 1857
Db 2475 -----ThrGlnArgValAsnAsnAlaGlnThr 2483
OY 1858 CATGTATTGATGAATGTACAGAAATTCCTTATGTTTGACAGATGGTTAAACAAA 1917
Db 2484 GlyLeuAsnGlyAspThrAsn-----LeuAlaThrAlaLys 2495
OY 1918 GAAGAGAATGAATAGTATTAAGAAACTGTTCAAAAAAAGAAATATACAGCATCG 1977
Db 2496 GlnGlnAlaLysAspAlaLeuArgGlnMetThrHisLeuSerAspAlaGlnLysGlnSer 2515
1978 TATATAGTATATTAATCTTTTGAAGCTTATTTTAAAGTTATGATTAACCTT 2037
2516 IleThrGlnLysIleAspSerAlaThrGln----- 2525
OY 2038 GACAAGATGACCAAAATGGAAGAACTTATGCAAAATATTAAGAAAAAATAATGAG 2097
Db 2526 -----ValThrGlyValGlnSerValLysAspAsn 2535
OY 2098 TTTTCCATTTGGAAATTAATAGGACTATTAGAAATGCAATAGAACTCTGTATGAT 2157
Db 2536 AlaThrAsnLeuAspAsnAlaMetAsnGlnLeuArgAsnSerIle---AlaAsnLysAsp 2554
OY 2158 CACTTAAAGAACT-----GCCAGATATGTAACAACTAATACAAAC 2202
Db 2555 AspValLysAlaSerGlnProTyrValAspAlaAspArgAspLysGlnAsnAlaTyrAsn 2574
OY 2203 GAAGCA-----TGTGAACATCCCATATATGCAACAACAAACCCGTGTGTAACCT 2253
Db 2575 ThrAlaValThrAsnAlaGlnAsnIleIleAsnAlaThrSerGlnProThrLeuAspPro 2594
OY 2254 CGTGAAGGACGCAACCCACTTAATAATAAAGAAATGCAACATCTTAAAGAGCT 2313
Db 2595 SerAlaValThrGlnAlaAlaAsnGlnValSerThr-----AsnLysThr 2609
OY 2314 GCATACGAGAGAACGAAATCGGTCTTCATTAATGGAAGAAAGCAGCAGCAAGGT 2373
Db 2610 AlaLeuAsnGlyAlaGlnAsnLeuAlaAsnLysGlnGlnThrThrAlaAsnIleAsn 2629
OY 2374 ATATATAACGCGGGGTGAGAAAGCACTCAAGCAATTTATGTAGATAATATGATA 2433
2630 -----GlnLeu 2631
OY 2434 AATCATTTCTAATCGT-----AATCTTGTTTTCAAATGAGCCATGT 2475
Db 2632 SerHisLeuAsnAsnAlaGlnLysGlnAspLeuAsnThrGlnValThrAsnAlaPro--- 2650
OY 2476 GATGCAAAAGCAGAGGATGTATACAAACA----- 2508,
Db 2651 ---AsnIleSerThrValAsnGlnValLysThrLysAlaGlnLysAlaMet 2659
OY 2509 ---AGATTTCGTAGAACTGAATGGAAGAGATGCCGGAACACATGCCGTAAGATCAC 2555
Db 2670 GlnArgLeuIleAsnGlyIleGln-----AspLysAspGlnValLysGlnSerVal 2686
OY 2566 GAAGATGTATTATTCCTCTAGAAAGACATATATGATACCAATTTGAGACATTTA 2625
Db 2687 AsnPheThrAspAlaAspProGlnLysGlnThrAlaTyrAsnAsnAlaValThrAlaAla 2706
OY 2626 CAAGAGATGATCAACCCCTTAATGTAATATTTGTATGATCTTTAGTTAATATTCCTTT 2685
Db 2707 GlnAsnIleIleAsnGlnAlaAsnGlyThrAsnAlaAsnGlnSerGlnValAlaAla 2726
OY 2686 TTGGGGGATGTTCTTATCATCAAAATATGACAAACAAAGATTAATACGAATGTATAAA 2745
2745
Db 2727 LeuSerThrValThrThr----- 2733
OY 2746 GAAGAATATTAATCAAGGCCCCAAGAGATTAAGTACGCCAAACACACAGCA----- 2799
Db 2734 ---LysGlnAlaLeuAsnGlnLysAspArgLysValThrAspAlaLysAsnAlaAsnGln 2752
OY 2800 ACTATCTGCGACCTATACGTTACAGTTTTCAGAT-----ATAGGTGATATA 2847
Db 2753 ThrLeuSerThrLeuAspAsnLeuAsnAsnAlaGlnLysGlnValAlaValThrGlnLysAsnIle 2772
OY 2848 ATTCAGAGAAGATCTCGGAAGAAAGAAACGGTACATCGTAAAGCTCAAGACATTTG 2907
Db 2773 AsnGlnAlaHisThrValAlaGlnValThr---GlnAlaIleGlnThrAlaGlnGlnLeu 2791
OY 2908 GAATCTGTTTGGTATATATACATTAAGTACCTCAACAGCAAA-----GGAAT 2955
Db 2792 AsnThrAlaMetGlyAsnLeuLysAsnSerLeuAsnAspLysAspThrThrLeuGlySer 2811
OY 2956 GATTAATATATGATGATGCCCCCAATATTTAAATTTGACGGAAATTCGTGGAA--- 3012
Db 2812 GlnAsnPheAlaAspAlaAspProGln-----LysLysAsnAlaTyrAsnGln 2827
OY 3012 ----- 3012
Db 2828 AlaValHisAsnAlaGlnAsnIleLeuAsnLysSerThrGlyThrAsnValProLysAsp 2847
OY 3013 -----GCTAATAGACCCAAAGTATGGAAGCCATGAATGTGATATGAA 3057
Db 2848 GlnValGlnAlaAlaMetAsnGlnValAsnAlaThrLysAlaAlaLeuAsnGlyThrGln 2867
OY 3058 TATTTGAAGCATTAATGCGGACCAACCAACAAAGTACTATTCGGATATAGTATGAT 3117
Db 2868 AsnLeuGlnLysAlaLysGlnHisAlaAsnThr-----AlaIleAspGlyLeuSerHis 2885
OY 3118 CATACACCATTTGAT---GATTATATCCACACAAATTAATGATGATGACGAATGGCA 3174
Db 2886 LeuThrAsnAlaGlnLysGlnAlaLeuLysGlnLeuValGlnGlnSerThrThrAlaAla 2905
OY 3175 GAATGTACTCAAGGTGCGAAAGAGAG-----TATGATAGTTGAAG 3219
Db 2906 GlnAlaGlnGlnAsnGlnGlnLysAlaAsnAsnValAspAlaAlaMetAspLysLeuArg 2925
OY 3220 GAGAACTGTAAGAGCTGTAAAGATTAAGTATGTCAGGCTGTCAAGAGAGTGTGT 3279
Db 2926 GlnSerIleAlaAsp-----AsnAlaThrThrLysGlnAsn--- 2937
OY 3280 ACAGTTGTACGAAGTGCACAGAAAGCTTATGATATATATATATATATATGATTAAG 3339
2938 -----GlnAsnTyrThrAspAlaSerGlnAsnLys 2947
OY 3340 AAGAACATGGAATATATATCATGATTAATCAAAAGATTAACATGACAGCA----- 3393
Db 2948 LysAspAlaTyrAsnAsnAlaValThrThrAlaGlnGlyIleIleAspGlnThrThrSer 2967
OY 3394 -----CAATGTCTGTACTAATTAAGTGTGAAGCTTCAGTACAGTCCCAAAAT 3444
Db 2968 ProThrLeuAspProThrValIleAsnGlnAlaAlaGlnValSerThrThrLysAsn 2987
OY 3445 CATATGACAGAGATTTATGATTTGCGAATTAATACCAAAAGTGGGCAAA 3504
Db 2988 AlaLeuAsnGlyAsn-----GlnAsnLeuGlnAlaAlaLysGlnGln---Ala 3002
OY 3505 AGTAATAAAGGTGACTAGTATGATGAAGTGTCTCATTTGACTATACACACAGATGAA 3564
Db 3003 SerGlnSerLeuGlySerLeuAspAsnLeuAsnAlaGlnLysGlnThrValThrAsp 3022
OY 3565 AATGTT---GGAGCATATCCTCATATACAGAAATTTTGAATGTGTCACTCAAAAT 3621
Db 3023 GlnIleAsnGlyAlaHisThrValAspGlnLysAsnGlnIleLysGlnAsnAlaGlnAsn 3042
OY 3622 GAGTTTGTATGAAGAAAGATGATGATGAAGATTAAGAAATATATGCTTATAGATATA 3661
3043 -----LeuAsnThrAlaMetGlyAsnLeuLysGlnAlaIleAlaAspLys 3057

QY 3682 CCACGAGCCATGATGTCGCTGCTGTTAAAGTGCATCCAGCAAGGCTACAG 3741
Db 3058 -----AspAlaThrIysAlaThr 3063
QY 3742 ATAAAAACGAAAAAACCGAGCAAGATACGGAATGTAAACACTG----- 3792
Db 3064 ValAsnPheThrAspAlaAspGlnAlaIleThrAlaThrAsnThrAlaThrAsn 3083
QY 3793 -----AATGATATACTTAAAGAAAACGATGGAAAG----- 3822
Db 3084 AlaGluAsnIleIleSerIysAlaAsnGlyGlyAsnAlaThrGlnAlaGluValGluGln 3103
QY 3823 -----AAACAAGTACAGACATTTGCATCCAAAAAGATAGTAATGCCGATGG 3876
Db 3104 AlaIleValGlnValAsnAlaAlaIleValGlnAlaLeuAsnGlyAsn----- 3118
QY 3877 CAATGGCGAAATATTAATTTAGTGGAGACCTCGTGTGTATGCCCCCTAGAACACA 3936
3119 -----AlaAsnValGlnHisAlaIleValAspGluAlaThrAlaLeu----- 3131
3937 AAGTTATGCGTACATTTCTGGCAATGATTAATGAATAAAAAATTAACATCACAAATT 3996
Db 3132 -----IleAsnSerSerAsnAspLeuAsnGlnAlaGlnIleIysAspAla 3145
QY 3997 AATTTAAAGAACTTTCAATCAATCTGCAGACAGCAAAACATTTCTCATGTATTTAT 4056
3146 -----LeuIysGln-----GlnValGlnAsnAlaThrValAlaGlyValAsnAsn 3161
QY 4057 TATTAAGTAAAGATGTCGAGCAAGATGCAATGCAATTAAGATTAAGAACGCAAAAT 4116
Db 3162 ValIysGlnThrAlaGlnIleLeuAsnAsnAlaMetThrGlnLeuIysGln----- 3178
QY 4117 CCTCCGCAATTTTGTAGATCATGTTCTACACATTTGGAGATTATGAGATTTTATTT 4176
Db 3178 ----- 3178
QY 4177 GGAACGATATATCAAAAGCTATGTCGAGCAAGTAAACTAAAGACAAATGATTC 4236
Db 3179 -----GlyIleAlaAspIysGlnIleThrIysAlaAspGly 3190
QY 4237 CTTTCAAAATGTGACCAAAATCTCTAATGCAAAAAACGCCAAGATGCTGACA 4296
Db 3191 AsnPheValAsnAlaAspProAspIysGlnAsnAlaThrAsn----- 3204
QY 4297 GAACATAGTCATGATATGAGAGCTATGCTATGTGCACTAATAAATTGGGCAAAA 4356
3205 -----GlnAlaValAlaIleValAlaGlnAlaLeuIleSerAlaThr 3217
QY 4357 AAAGATGATTTTACCAAAACCTACGGTTACACACAGCTCAAAATTGTCGACAAAC 4413
3218 ProAsp-ValValAlaValThrProSerGluIleThrAlaAlaLeuAsnIleValThrGlnAla 3237
QY 4414 ACCACTTGGAGAAATTTGCCAAGACCCAGTTTTCAGATGCTAACCGAATGGTAC 4473
3237 A-----LysAsnAsp-----LeuAsnGlyAs 3244
QY 4474 GACGACTATGCTATACACGACAAAAATTTGAAGATGTCGAGCAAAATGTAACTA 4533
3244 nThrAsnLeuAlaThrAlaIleValGlnAsnVal-----GlnHisAlaIleAspG 3260
QY 4534 AATGACCAATTAAGTGTATACAG----- 4558
3260 nLeuProAsnLeuAsnGlnAlaGlnThrAspGluIleThrIleThrGlnAlaThr 3280
QY 4559 -----AATGATATAGAAATGCGAGGATAGCTAGCTTAATAT-----ATGAA 4598
3280 nLeuValProAsnValAsnAlaIleGlnIleAlaAlaThrThrLeuAsnAspAlaMetThr 3300
QY 4599 AAAAAAAGAGTGCATTC-----CAAGATTAATATTACAA 4637
3300 nGlnLeuIysGlnGlyIleAlaAsnIysAlaGlnIleIleIysGlySerGluAsnThrIleAs 3320

QY 4638 GCATGACCCGACAAAA----- 4656
Db 3320 PalaAspThrAspIysGlnThrAlaThrAspAsnAlaValThrIysAlaGluGluLeu 3340
QY 4657 -----AGATTGATAGACACACATTTGGTGTATGCTACAGACTA 4697
Db 3340 nIysGlnThrThrAsnProThrMetAspProAsnThrIleGlnGlnAlaLeuThrIysVa 3360
QY 4698 TACTGACGAAT----- 4710
Db 3360 IAsnAspThrAsnGlnAlaLeuAsnGlyAsnGlnIysLeuAlaAspAlaIysGlnAspAl 3380
QY 4711 -----GCAACGATTTACTTGAAC-----AGCAAAATTTACTGC 4742
Db 3380 nIysThrThrLeuGlyThrLeuAsnHisIleAsnAspAlaGlnIysGlnAlaLeuThrThr 3400
QY 4743 TAGTGTGTGATTAAGCCCTGGAAGTCCCTGTG-----GTACAAAGAAATATCAAT 4796
3400 nGlnValGluGlnAlaProAspIleAlaThrValAsnAsnValIysGlnAsnAlaGlnAs 3420
QY 4797 GTTAGAA-----AAACAGCTTCTATGATGCCGACAAACATTTGTG 4838
3420 nLeuAsnAsnAlaMetThrAsnLeuAsnAsnAlaLeuGlnAspIysThrGlnThrLeuAs 3440
QY 4839 GTGCACAAAATTTATTTGAATATGAC-----GACAAATATATCTACATTTGAG 4886
Db 3440 nSerIleAsnPheThrAspAlaAspGlnAlaIleIysIysAspAlaIleThrAsnAlaValIse 3460
QY 4887 TAAAGTAAAGTCAAAAGATTTAGTAAAGGCAACACAGTGCATTAAGTGGCAAA 4946
3460 r-----HisAlaGluGlyIleLeuSerIysAlaAsn----- 3470
QY 4947 CAAGGTCCTAATTAACATACTTAACATTAACATTAACATTAACATTAACATTAAC 5006
3471 -----GlySerAsnAlaSerGlnThr-----GluValGlnGlnAlaIle 3483
QY 5007 TCGTCGACTACATATATGTTTTCATGCAATGCAATTAATACAGATCCAGAACTTA 5066
Db 3483 nGlnThrValValAsnGlnAlaIleIysGlnAlaLeuAsnGlyAsnAspValGlnThrAlaIle 3503
QY 5067 AGATGAAATATGGGTTCCGAAAAAGATTTGATGGAAGTGGGCGCAAGGAGTACAAATT 5126
3503 sAsp-----AlaAlaIysGlnValIleThrAsnAlaAsnAspLe 3516
QY 5127 GGCTCAATACACAAAGAA-----AAAAAGAAAAAGAAATTAACATCGCGATGC 5180
3516 nAsnGlnAlaGlnIlyAspAlaLeuIysGlnIleValAspAlaIleThr----- 3533
QY 5181 GCACAAATATTTTATGAGCTCCCGCTTGTACTGCTATGAATATATGTTTATGATTT 5240
3534 -----ValAlaAsnValAlaAsnThrIleIysGlnThrAlaGlnAspLe 3547
QY 5241 AAGAGATATTAATCTGATGATGATTAATTTGGAAGATGGAAGAAACAGCGAGGAAA 5300
3547 nAsnGlnAlaMetThrGlnLeuIysGlnIlyIleAlaAspIysAspGlnThrIysAlaAs 3567
QY 5301 TTTGAAGAAATA-----TTTAACAAAAATGGAACATGCTTGCAAA 5342
Db 3567 nGlyAsnPheValAsnAlaAspThrAspIysGlnAsnAlaIleThrAsnAlaValAlaIle 3587
QY 5343 AGGAAGTATAGTACTACAGCAAAATCCCGTACTGCGCGCAAAATTTTTCGACACA 5402
3587 sAlaGluGlnIleIleSerGlyThrProAsnAla-----AsnVa 3600
QY 5403 AATAAGCAATGTGTGGAAGCAATG-----ATATCGGGGTACAAACGTTAGAGATGA 5459
3600 nAspProGlnGlnValAlaGlnAlaLeuGlnGlnValAsnGlnAlaIysGlyAspLeuAs 3620
QY 5460 TCGAAATAGTGAATATAGTCAAGAAAGATGATGAAGATCTAAAAAATGTGTTCTGTACC 5519
3620 nGlyAsnHisAsnLeuGlnValAlaIleIysAspAsnAlaAsnThrAlaIleAspGlnLeuThr 3640
QY 5520 TTCAGATGATGATTTATCTATGTGGGAAAAATCGCGATGAAGGTACTGCTATCAGTTTCT 5579

OY	7510	-----	CCATATGAGCTGTCACAAAGAAAGAAAG	7558
Db	4245	rLeuGlnAlaLeuAsnGlyAspHisAsnLeuGlnValAlaLysThrAlaSerAlaThrGlnAl	4265	
OY	7539	TAAAAATGTTGGACTAAATATATATCTTCGATGCTCAAGAAACCTTAAGACCTGAT--	7596	
Db	4265	AlaAspAlaLeuThrSerLeuAsn-----AspProGlnLysThrAlaLeuLysAspG	4283	
OY	7596	-----	7596	
Db	4283	nValThrAlaAlaThrLeuValThrAlaValHisGlnIleGlnGlnAsnAlaAsnThrLe	4303	
OY	7596	-----	7596	
Db	4303	uasnGlnAlaMetHisGlyLeuLysArgInserIleGlnAspAsnAlaAlaThrLysAlaAs	4323	
OY	7597	----AAATATATAGAAAGAAAGAACCCCTTGCAAAATAGAAAGAAATCGTTTAAG--	7650	
Db	4323	nSerLysTyrIleAsnGlnLysPrlnProGlnGlnHisnGntYrAspGlnAlaValGlnAl	4343	
OY	7651	-----GTGATTTATGAAATGAAATGTTACAAAAATCGTAAGTTCTATACGA	7697	
Db	4343	AlaAsnAsnIleIleAsnGlnGlnThrAlaThrLeuAspAsnAsnAlaIleAsnGlnAl	4363	
OY	7698	GAIAAAAGAGTATGCTGCTACCTCCAGAGAGACATATGCTCTTAAGAAATTATAGATGA	7757	
Db	4363	AlaIleThrValAsnThrThrLysAlaAlaLeuHis-----GlyAs	4377	
OY	7758	AATTAATTAATGAAAGACTTAAGAT-----	7782	
Db	4377	PValLysLeuGlnHisnAspLysAspHisAlaLysGlnThrValSerGlnLeuAlaHisLe	4397	
OY	7783	-----AGTAATATATCTCCCTAAATAGCTTCGTCGAACCTGC	7817	
Db	4397	uasnAsnAlaGlnLysHisMetGlnAspThrLeuIleAspSerGlnThrThrAlaThrAl	4417	
OY	7818	ACGAAATAGAGAAATACATATATTAATAAAACTTCACACTCAGAGAACGGGTGGCAATGA	7877	
Db	4417	avalLysGlnAspLeu-----ThrGlnAlaGlnAlaLeuAs	4429	
OY	7878	TCCAAATATGTGATACTATGAAATATAGTTGCTGCTGATCTGGGTGACATAGTTAGAGAAC	7937	
Db	4429	PglLeuMetLysPrlaLeuGlnGlnSerIleAlaAsp---LysAspAlaThrAlaGlnAlaSe	4448	
OY	7938	AGATATGTTACGAATTCGTGCTTACTTACCTCCCGTAGAAATTAATATATAGCTTTT	7997	
Db	4448	rSer-----	4449	
OY	7998	TGAATCATATATGGAATGAGAAATGAGAAATTAATAAGGTGAAATTAATATCAACGATGT	8057	
Db	4450	AlaTyrValAsnAlaGlu-----ProAsnLysLysGlnSerTyrAspGlnAlaVala	4466	
OY	8058	ACAAACGTTTCGTTGCTGCTGTGGGATGCTAATTAAGAAAGATATTTGGAAGCAATGAC	8117	
Db	4466	IglHisAlaGlnSerIleIleAlaGlyLeuAsnAsnProThrIleAsnLysGlyAsnVa	4486	
OY	8118	GTGCAAGACACACAGAGATGCAAAACTTTTGAAGAAAGCAAAAGATGATTTGAACG	8177	
Db	4486	IserSerAlaThrGln---AlaValIleSerLysAsnAlaLeuAspArgValGlnAr	4505	
OY	8178	CATAACATTATACAGATGATG-----TGTGACATTAAGACAGAT--	8217	
Db	4505	g-----LeuAlaGlnAspLysGlnThrAlaGlyAsnSerLeuAsnHisLeuAspArgIle	4523	
OY	8218	-----CCACCTGTTGATGATTTATATACCTCAACAGGTTTCGATGATGACTGAATGCTGCA	8273	
Db	4523	uThrPrlaGlnGlnGlnAlaLeuGlnLysnGlnIleAsnAlaAlaThrThrArgGlyG	4543	
OY	8274	ATATATATATGTTAA-----CCACGTGATGCAAGAAATTTGAAATATTTAAATA	8318	
Db	4543	uValAlaGlnLysLeuThrGlnAlaGlnAlaLeuAsnGlnAlaMetGlnAlaLeuAlaValAs	4563	

OY	8319	ATCATGTCATCATGTAAACATCTGCAGATGCACAAAGTAAATGATTTATGATGAATTAAGT	8378
Dd	4563	nserTlleglnasprcInclnThrglnalaglyserLyspheIleasnglAspLysPr	4563
OY	8379	TGAACAGTGTAAACGAGATGTCAGAGATTTAAAAATTTTGTCTTAAATGGAAATCTCT	8438
Dd	4583	oGln-----	LyAspAl 4587
OY	8439	ATTGCAATTCATCAATCAAAATAAATACAAAGATTTGTATGAAACCAATATATACAAAAAT	8498
Dd	4587	aTyglnAlaIaIaValglInasnaIaLysasrLeuIleasngIn-----Thrasnas	4604
OY	8459	CTGTACTATATCATGTCTGAAAAATTTTGTACAAAAGTTGAAACCTTTTAAAGATG	8558
Dd	4604	prroThreLysaspLysAlaGln-----ValgluGlnLeuThrgln-----	4617
OY	8559	TTTGTGTGAGAGCTTTTCTGATATCTCTGTAAGAACAGTAAGTGTGATTAATAAA-	8616
Dd	4618	-AlaValasnglnAlaLysasrpsnLeuIleasglLysaspGlnLysLeuAlaAspLysGl	4637
OY	8617	-----TTTAATGAAAAATGATGTCTTCATATATACGAATATATGCTTT	8660
Dd	4637	nHsAlaValaThrasprLeuasnGlnLeuAsnGlnLysAsnAspProglInarglnAlaIe	4657
OY	8661	CGAA-----	GAACACCAAAAAGTTATAAAGAAC
Dd	4657	uGlnserGlnIleasnasnaIaIaThrargLysuValaIaGlnLysLeuAlaGlnAl	4677
OY	8691	TTTGCAGTTGTACACTACTACTCTTCTAAGAAATCCATGGATATTTGTCCACAGAT-----CA	8744
Dd	4677	AlaValaLeuAsp-----GlnAlaMetGlnAlaLeuArgAsnSerTleglnAspGlnGlnI	4696
OY	8745	AAACAAAGATGATGATGAAGAATTAACAACTTTTACCTTCTGCTGCAAGAAATGATATGA	8804
Dd	4696	nThrgLyserGlyserLyspheIle-----AsnGlnAspLys	4708
OY	8805	TAAATATCTTATATATTGGAACGCATACCTTGTCTTAATAGTTCAGAT-----	8833
Dd	4708	sProglInLysAspAlaTyrglnAla--AlaValglInasnaIaLysAspLeuIleasngl	4727
OY	8854	-GATACAAAGCTGATATGATTTCTCTCAACAAACAAACATTTATGTACAAAGCTATGAC	8912
Dd	4727	nThrgLysnPrroThreLysaspLysSerGlnValaIaGlnLeuThrglnAlaValaThTh	4747
OY	8913	TGCATATATATATGAAGAAAGTCAATAAGCAAAATTTTAAAA-----	8952
Dd	4747	rAlaLysAspAsnLeuIleasglLysaspGlnLysLeuAlaIaArgAspGlnGlnAlaValaTh	4767
OY	8953	-----	AAAAAATCTTACTTCTGCTTT
Dd	4767	rThValasnaIaLeuProAsnLeuAsnHsIaGlnGlnGlnAlaIaLeuThraspAlaI	4787
OY	8976	CAGTCACAGCAAAATG-----TTAGTCAAAAATATTAATGCGAAGAACAGTGT--	9024
Dd	4787	eAsnAlaIaIaPrroThrArgThrglnuValaIaGlnHsValaIaGlnThrAlaThrgLyuAs	4807
OY	9025	-----TGCCTTTGAGCCATGAAA-----	9042
Dd	4807	rHsIaIaMetGlnuThreLysAsnLysValaAspGlnuValasnaThraspLysAlaGlnPr	4827
OY	9043	-----TATAGTTATGCA-----	9054
Dd	4837	oAsnTyThrgLysuIaIaSerThraspLysLysGlnuValaValaAspGlnAlaIaLeuGlnAlaI	4847
OY	9055	-----GATTATTTCCGATATATTAATTAAGCAATGATATGATGACACCTCTATTCTCGAAA	9110
Dd	4847	aGlnSerTleThraspPrroThrasngLysSerAsnAlaAsnLysAspAlaValaAspGlnuVa	4867
OY	9111	AATTAATAAAATATTTATTAACATCAATCAATGAAGCAACCAAAATTCG-----	9156
Dd	4867	lLeuThrLysLeuGlnLyuLysGlnuAsnGlnuLeuAsnGlnLysAsnGlnuArgValaIaGlnAl	4887
OY	9157	AAAAATGCTGGGAAAAATATAGAGCTCAGATATGACCGCTATATGTATGTGATATTA	9215

Db	5204	ntgttlaasnlaasprsaenleuysalaasnltgtaspsnalailealaasnala	5224
Qy	10117	-----TTATCTCTCTGATATTATCTGATCTGCAGAAAGTAGATGAGA	10161
Db	5224	ahiscldeuabrslysaingniyaasnalailealaysalagialaglnglnleu	5244
Qy	10161	ATTGGATTTTAATGATTA	10194
Db	5244	sglnasilelleasprlacllnasnalaleuasnlglyaspglnasnldeuallaasnaly	5264
Qy	10194	TCTAAATATTAACCATTTAGACAGTGA	10246
Db	5264	sasrlysalasnalaphavealasnserleuasnnglyleuasnnglnglnlnasple	5284
Qy	10246	-----ATACCAAGTATATACACCAAGTAATGATACACCAAGTAGAATAG	10293
Db	5284	ualanileylasalaileasnynlaasprhvalaserasvaltghraspillevalasn	5304
Qy	10293	ATTATTTGAT-----GATGAAGGAATGAACTGAAACATGATTTGTATCTCAATATT	1034
Db	5304	nglnlleaspleasnsprlamecsluthrleuysnhs--leuValaaspsnngluil	5323
Qy	10347	ACCAATACAGAACCAATATATATATACAAAGTCAGATATTCOAATGATACAGAAC	1040
Db	5323	eprodnsalaglnglnthrValasnrycllnasnlasp-----Aspasnal	5339
Qy	10407	TAAATCTTTATATCTGATTAATCTCGAAGAAAACCTTTATATTATCTATTCATGATAG	10461
Db	5339	alvsthrasnpheasprsp-----	5345
Qy	10467	GGATTTATATACTGGGAAGAAATTAGTTAAATTTAATTTGGAGTCAATATAAA	10524
Db	5346	-----AlaLysargleualasnthrleuasnseraspsntrhaan--	5360
Qy	10527	TGATATTCOAATGAAATGCTAGAATGATCTTATAGAGGTATAGATATTAAATGTC	10586
Db	5361	-----ValasnAspilleasnnglyalallelnalavayalasnspal	5374
Qy	10587	ACTAGAGTGCCTAAAC	10602
Db	5374	allelnasnldeasn	5379

```

RESULT 14
: US-09-919-172-29
: Sequence 29, Application US/09919172
: Patent No. US20020119463A1
: GENERAL INFORMATION:
: APPLICANT: Fairis, Mary
: APPLICANT: Turner, Christopher M.
: TITLE OF INVENTION: PROSTATE CANCER MARKERS
: FILE REFERENCE: PA-0036 US
: CURRENT APPLICATION NUMBER: US/09/919,172
: CURRENT FILING DATE: 2001-07-30
: PRIOR APPLICATION NUMBER: 60/222,469
: PRIOR FILING DATE: 2000-07-28
: NUMBER OF SEQ ID NOS: 102
: SOFTWARE: PERL Program
: SEQ ID NO 29
: LENGTH: 2125
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc.feature
: OTHER INFORMATION: Incyte ID No. US20020119463A1 3774181CDT
US-09-919-172-29

Alignment Scores:
Pred. No.:
Score: 1.59e-10
Percent Similarity: 291.00
Best Local Similarity: 32.60%
Query Match: 18.13%
DB: 1.50%
10
Gaps: 122
Matches: 2125
Conservative: 466
Mismatches: 364
Indels: 837
Gaps: 859
122

```


US-10-087-013-1 (1-10628) x US-09-919-172-29 (1-2125)

OY 3718 GGATGAAACCCAGAGGTACAGATTAACGAAAAAAGCGAAGAAAGATACG 3777
||| ||| :|||:||||
Db 19 GlyLysAlaLysThrIleLeuLeuLys-----ProArgSer 32

OY 3778 GAATGTAAACAGTAAATGATATA--CTTAAAGAAACGATGAAGAAACAGTAGAA 3834
:|||| :|||
Db 33 AspCysProLeuLysThrSerIleProIleLysAlaIleCysAspTyrArgLniLeu 52

OY 3835 -----GATTGTCAATCCAAAAAGATAGTATGATATCCCAT 3873
:|||| :|||
Db 53 IleThrIleTyrLysAspAspIleCysValLeuAlaAsnSerHis---ArgAlaLys 71

OY 3874 TGGCAA-----TGGGAAATATTAATTTAGTAGAAGACCCCTGCTGTATG 3921
||| ||| :|||:||||
Db 72 ThrValValIleSerProThrIleLysnGluAlaMetVal-----ProSerValCysPhe 89

3922 -----CCCCCTAGAACACAAAGTATGCTACATTTCTTGGCAAT-----GAT 3966
||||| :|||:||||
90 ThrValProProAsnLysGluAlaVal-----AspLeuAlaAsnArgLniLeu 106

OY 3967 AATGAAATTAATAAATA-----CAATCAAGTTATTTAAAGAAAGCT 4011
:|||| :|||
Db 107 GlnGlnTyrGlnAsnValLeuThrIleThrHisGlnSerHisIleAsnMetLys----- 124

OY 4012 TTTCAATCAATCTGCAGACAGACAAACATTTCTTCATGATATTTATTAAGTAAGAT 4071
:|||| :|||
Db 125 -----SerValValSerThrHisTyrIleLeu----- 133

OY 4072 GGTGAAGAAATCAACTGCATAA-----GAATTAAGAAAGGC 4110
||||| :|||:||||
Db 134 -----AsnGluIleAspArgIleArgAlaSerAsnValAlaSerIleLysThr--- 149

OY 4111 AAAATTCGCCCGCATTTTGATCCAGTCTTACACATTTTGATGATATATAGATTTT 4170
:|||| :|||
Db 150 MetLeuProGlyGlnHisGlnGlnValLeuSerAsnLeuGlnSerArgPheGlnAspPhe 169

OY 4171 TTA-----TTTGAACAGATATATCAAAAGCTCATGCT 4203
||| ||| :|||:||||
Db 170 LeuGluAspSerGlnGlnSerGlnValAlaPheSerGlySerAspIleThrGlnLeuGluLys 189

OY 4204 GAGGAGCTAACTAAAGACCAAAATAGATTCCTTTTCAAAATGGTACCAAAATCT 4263
||| :|||:||||
Db 190 GlnValAsnValCysLysGlnTyrGlnGlnLeuLeuLysSerAlaGlnArgGlnGln 209

OY 4264 CCTATGAAAAACACGCCAAGATGTGACAGAAACATATGCTAT----- 4308
:|||| :|||
210 GlnGlnGlnSerValTyrAsnLeuTyrIleSerGluValArgAsnIleArgLeuArgLeu 229

OY 4309 GAGATATGGAGCTATGCTATGCTACATAGTAAATTTGGGCAAAAGAAATGATTTT 4368
||| ||| :|||:||||
Db 230 GluAsnGlyGluAspArgLeuIleArgGlnIleArgThrProLeuGlnLysArgAspLeu 249

OY 4369 ACCGAAACTACGCTTACACACAGCTCAATTTAGTACAAAAGCACCTTTGAGGAA 4428
:|||| :|||
Db 250 HisGlnSerVal----- 253

OY 4429 TTTGCCAAACGACCCAGTTTTCAGATGCTAACCGAATGTAGACAGCATATGCTAT 4488
:|||| :|||
Db 254 -----PheArgIle 256

OY 4489 ACACGACAAAATATTTGAAGATGTGACGAAAAATGTAAATGACCAATTTGAAG 4548
||| ||| :|||:||||
Db 257 ThrGlnGlnGlnLysLeuLysLysGlnLeuGlnArgLeuLys-----AspSerLeuGly 274

OY 4549 TGTGATACGATGTATATAGAAATCCGAGGCTACGTTAAATATATGAAAAA 4608
||| ||| :|||:||||
Db 275 ThrIleThr-----AsnLysCysGlnGlnPhePheSerGlnAlaIleAlaSerSer 291

OY 4609 GAGTGAATTCAC-----CAAGTAAATATTTACAGATGAAACGCAAAAAA 4656
:|||| :|||:||||

Db 292 Ser---ValProThrLeuArgSerGln-LeuAsnValValLeuGlnAsnMetAsnGlnVal 310

OY 4657 AGATTCGATAGACAAACACATTTGCTAATGCTTACAGACTATACT-----GG 4703
||||| :|||:||||
Db 310 LysSerMetSerSerThrTyrIleAspLysLysThrValAsnLeuValLeuLysAs 330

OY 4704 AACGAATGCACAGATTTACTTGAACAGAAATTT---ACTGCTAGTTGTGTGATTAAGCC 4760
||| ||| :|||:||||
Db 330 nThrGlnAlaIleGlnAlaLeuValLysLeuTyrGlnThrLysLeuCysGlnGlnGln 350

OY 4761 TGGAAAGTCCCTGTGTGTCAAAAGAAATATACAAATTTGTA----- 4800
||| ||| :|||:||||
Db 350 aValIleAlaAsp-----LysAsnAsnIleGlnLysnLeuIleSerThrLeuLysGlnThr 368

OY 4801 -----GAAACAGGCTACATATGATGCGCAACAAACATTTGGGGTG 4841
:|||| :|||
Db 368 PArgSerGlnValAspGlnLysArgGlnValAlaPheHisAlaLeuGlnAspGlnLeuGlnLys 388

OY 4842 CACAAATTTATTTGAATGACGACAAATATATCAATTTGAGTAAGAT----- 4893
||| ||| :|||:||||
Db 388 AlaLysAlaIle-----SerAspLysMetPheLysThrTyrLysGlnArgAspLeuAspPhe 407

OY 4894 -----AAGTCGAAAGATTAAGGAGCGCAACACAGGTGCTATTAA 4937
:|||| :|||
Db 407 eAspThrHisLysGlnLysAlaAspGlnLeuValGln-----Ar 420

OY 4938 GTGGCAAAACAAAGGCTCAATATATCAATATCAATTAAGTAAGTAAGTGAATGCT 4997
||||| ||| :|||:||||
Db 420 gTrpGlnAsnValHisValGlnIleAspAsnArgLeuArgAspLeu---GlnGlyIleLeu 439

OY 4998 TTTTCTCTGCTCGCATGATATGTTTCAATGATGATGCAATTAATACAGATCC 5057
:|||| :|||
Db 439 LysSerLeuLysTyrTyrArgAspThrTyrHisProLeu-----AspAs 454

OY 5058 AGAAGTAAAGTGAATAAGGCTGGGAAAGATTAAGTAAGTAAGTGGCGCAACGAAAG 5117
:|||| :|||
Db 454 pTrpIleGlnGlnValGlnThrThrGlnArgLysIleGlnGlnAsnGlnProGlnAsnSe 474

OY 5118 GTCAATTTGGGCTCAATATCAATCAAAAGAAAAA-----GAAAAAGA 5159
||| ||| :|||:||||
Db 474 LysThrLeuAlaIleThrGlnLeuAsnGlnGlnIleLysMetLeuValSerGlnIleGlnMetLys 494

OY 5160 GAAATTAATAACGCTGGATGCCACAAATATTTCTATGAGGTCGCCCTTGTAGTCTAT 5219
:|||| :|||
Db 494 sGlnSerLysMetAspGlnCysGlnLysTyr-----AlaGln 506

OY 5220 GAAATATAGTTT-----TATGATTTAAGCATATATCTAGTATGATTA 5267
:|||| :|||
Db 506 uGlnTyrSerAlaThrValLysAspTyrGlnLeuGlnThrMetThrTyrArgAla----- 524

OY 5268 TTTGGAAGATGAAAAACAAAGACGAGGAAATTTGAAGAAATATTTAAACAAAAATGG 5327
:|||| :|||:||||
Db 525 -MetValAspSerGlnGlnLysSerProValLysArgArgMetGlnSerSerAlaAs 544

OY 5328 AACATCAGTTGGCAAGGAGATGATCTACAGGAAATCCCGGTAGTACTGCGGAAA 5387
:|||| :|||
Db 544 PleuIleIleGlnGlnPheMetAspLeuArgThr----- 555

OY 5388 ATTTTCTGAACGAAATTAAGAAATGTGTGGAACGCAATGATATGCGGGTACAAAGC 5447
:|||| :|||
Db 556 -----ArgTyrThrAlaLeuValThrLeuMetThrGlnThrIleLys 569

OY 5448 TGTAGAGATGATGAAATAGTGAATAGTCAGAA-----AGTGAAGAAAGATCTAAA 5501
:|||| :|||:||||
Db 569 s-----PheAlaGlyAspSerLeuLysArgLeuGlnGlnGlnLys 584

OY 5502 AAAATGTGTTCTGTACCTTCAAGATGATATATCTATGGGAAAAATCCGAGTAAAGC 5561
:|||| :|||
Db 584 sArgCys-----LysGlnThrSerGlnIle 592

OY 5562 TACTGCTATCAGATTCTTCGATGCTTGCAGAAATGGGCTGAAGATTTTGCAAACATTA 5621
||||| ||| :|||:||||
Db 592 sGlnLysArgSerAspLeu-----LeuGlnArgGlnLys 603

```

QY 5622 AGAAAAGAAATGGAGAAATGAGAGGGCGGTAAATGATTAATCTGTGGTGAATAATGA 5681
Db 603 salatThrValLeuGlu----- 608
QY 5682 AGATTAAGAAAGAAATGTCACATGCGGTGACCAATATATAAAATTTATAGTAGTG 5741
Db 609 -----AsnSerLysLeuThrGlyLysIleSerGluLeuGlnArgMetValAlaGluLe 626
QY 5742 GAAACCAACAGTATGAAAAACAAATCAAAATATGTGAGAAATTAACCAAAATATATTC 5801
Db 626 uLyLeuGln-----LysSerTrpValGluG 635
QY 5802 CGACATCTCTGTGGCAAAAGATGACAGAGCGCTCGCAATATTTAGCAAAATTAATGA 5861
Db 635 uGluLeuProLysValArgGluAla-----AlaGluAsnGluLeuArgLysGlnGlnArg 653
QY 5862 AAAAAATTTGTAAATAATTAAGTGGAGATTTGTAATTAAGTATGAAAGATGTGCAC 5921
653 gAsnVal-----GluAspIleSerLe 660
5922 ACAGCGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5981
Db 660 uGlnLysIleArg-----Al 665
QY 5982 AGAAGTTGAAGAAAGTATATTTGTAATGTCACAGAGGTCCACACGCTGACGAGGGA 6041
Db 665 agLysSerGluAlaLys-----GlnTrpArgTrpG 675
QY 6042 AACACCGTACACAGCGGTATCACTGATATCAAAAGAGCGGATGCAAAAAGAGCGAA 6101
Db 675 uLeuGlu-----ThrIleValArgGluLysGluAlaAlaGluArgLysLeu-- 690
QY 6102 AACAGCGCGCCCTACAAAAGCGCAAAAGTGAAGAAATCTAACACAGAAATGCGAGC 6161
Db 691 -----GluArgValArgGlnLeuThrIleGluAlaGluAl 702
QY 6162 ACAAAACGACACCGCAGCAGCAGCA-----CAACAAC 6194
Db 702 a-----LysArgAlaAlaValGluGluAsnLeuAsnPheArgAsnGlnLe 718
QY 6195 ACGAAAACGACATCAACAGCAACACAGCAACAGAAATCTGACGTGGCAGCAATGGTAAAGC 6254
Db 718 uGluGluAsnThrPheThrArgArgThrLeuGluAspHisLeu----- 732
QY 6255 CATCTTTGCAATTAACAGATAGCAGGGGTGCAATAGAGGTTGTAATCCAAAAACGTA 6314
Db 732 ----- 732
6315 TGGACAATATCTTAATAGGGGTGTATGTAGTAACTTAAGAAATGAATAATGCGAT 6374
Db 732 ----- 732
QY 6375 ATGTATGCTCTAGGAAAAAAATTAATGTAATAATTAATTAACAATAT----- 6423
Db 733 -----LysAlaGlyLysAspLeuSerLeuAsnAspLeuGlnGlnLysAsnLys 748
QY 6424 -TTAAATTAATGAACCTGAAATTAAGCGTACAGATGATTAATAAGCCTTTATTAATG 6482
Db 748 sleuMetGluLeuLeuArgArgLysArgAspAsnGlu--GluGluLeuLeuLysLeuI 767
QY 6483 TGCAGCAATAGAACTCAATTTTGTGGTTAAATATATATTAATTAATAATCCTGACAGCA 6542
Db 767 eLyGluMetGluLysAspLeuAlaPheGlnLysGlnValAlaGluLysGlnLeuLysG 787
QY 6543 AAATGAATTCGAAAAATGGAACAATTCAGATGATTAATAAGATTAATGATTAATACATA 6602
Db 787 uLyGlnLysIleGluLeuGluAlaArgArgLysIleThrGluIleGlnLys----- 804
QY 6603 TGGTGAATTAATAAGATATGTTTTTGAAGCTGATTTCTAATGATTAATAATAATTAATAC 6662
Db 805 -----ThrCysArgGluAsnAlaLeuProValCysPr 815
QY 6663 TGTAAAC-----AATAGTGAACACCATCTTCATGAATAATTAATAGAA 6707
Db 815 oLleThrGlnAlaThrSerCysArgAlaValThrGlyLeuGlnGlnLysAspLysG 835
QY 6708 AAAACGAGAT-----AAAAAAGA 6728
Db 835 nLySAlaGluGluLeuLysGlnGlnValAspLysLeuThrAlaAlaAsnArgLysAlaG 855
QY 6729 TGAAGAAATTAACATAAATATTTTGGAGAAAAATTAATAATTTATTTGGAAAGACAT 6788
Db 855 uGlnAspMetArgGluLeuLeuThrTrpGluLeuAsn----- 866
QY 6789 AATAGATTAATCTTTCATCTGCAGACAGCAAAACGAAAGAAAAATTAAGATTAATTA 6848
Db 867 -----AlaLeuGlnLeuGluLysThrSerSerGluGluLysAlaArgLys----- 881
QY 6849 CCAGTACAAATGACATGACCAAACTGACCCCTCCCTGAAAGAGTTTGTAATAAGCCCCCA 6908
Db 882 -----LeuLysAspLysLeuAspLysLeuThrAsnAsnThrLeuArg 894
QY 6909 ATTTTGAAGATGTTTCACAGAAATGGCAGAGAAATTTTGTAAATGAAGAGAACAGATT 6968
Db 894 gCysLeuLysLeuGluLeuGlu-----ArgLysAspGlnAl 906
QY 6969 GTTAAAT-----TTGAGCGCGGCTGTAAAGAAATATAGTGTAAATGG 7010
Db 906 agLysLysGlyTrpSerGlnGlnLeuArgGluLeuGlyArgGlnLeuAsnGlnThrThrG 926
QY 7011 TAGTAATGACGTAAGACACAGCAAGATGTCAGAGCGCTGTGTAACATATCAAAATTTAT 7070
Db 926 yLyAlaGluGluAlaMetGlnGluAlaSerAsp-----Le 938
QY 7071 TAAAGAGCGAAATCTGAATTAAGAGCAAAAGCAAAAGTGAAGAAATGAAGATGAG 7130
Db 938 uLysLysIleLysArgAsnTrpGlnLeuGlnLeuGlnSerLeuAsnHisGluLys--G 957
QY 7131 CAAAAAGTAAAGATTAATCTTCTACTGAAGAGACATGAGAGCAACATGATGCTCA 7190
Db 957 yLys-----LeuGlnAlaArgGluValAspArgIleThrArgAlaAsn 970
QY 7191 T-----GAAATTTTAAACATGAATAATTAAGAAATTAATGTCGCA 7229
Db 970 sAlaValAlaGluLysAsnIleGlnHisLeuAsnSerGlnIleHisSerPheArgAsp 990
QY 7230 TAAAGATTTGTTCT-----TGTATGCAAAAACCT-----TCTTCAACAAT 7268
Db 990 uLysGluLeuGluArgLeuGlnIleCysGlnArgLysSerAspHisLeuLysGluGlnPh 1010
QY 7269 ACCAAAAACACACACAA-----TCACAATCATCGATGTAATGATAT 7313
Db 1010 eGluLysSerHisGluGlnLeuLeuGlnAsnIleLysAlaGluLysGluAsnAspLys 1030
QY 7314 GCCAGATCGCTGATTAATGTTCTCTGAAGAATTTTAACAAGTGT--GAGTGTCTGAAT 7370
Db 1030 sIleGlnAlaGluLeuAsn-----GluGluLeuGluLysSerAsnGluCysAlaGluMe 1047
QY 7371 TTCAAAAAGGCGATCTATGATTCATCAAAAATAATTAAGTGAATAATTAATCTGAATGA 7430
Db 1047 tLeuLysGlnLysValGluGluLeuThrArgGlnAsnAsnGlnIleThrLysLeuMetCysG 1067
QY 7431 TTCTGTAGAGAAACGCA----- 7449
Db 1067 nArgIleGlnAlaGluSerGluAsnIleValLeuGluLysGlnThrIleGlnArgCys 1087
QY 7450 -----TATTA 7454
Db 1087 sGluAlaLeuLysIleGlnAlaAspGlyPheLysAspGlnLeuArgSerThrAsnGluH 1107
QY 7455 TTTATCTTAAGAGAGCAAAATTAATATGATTAATCTTGAAGCAAAATTAATATCT-- 7512
Db 1107 sleuHisLysGlnThrLysThrGluGlnAspPheGlnArgLysIleLysCysLeuGluG 1127
QY 7513 -----ATTGAGTCTTAACAAAGAAAG--GAAAGTAAAAA 7544

```

Db	1127	uaspleuAlalysserclinsnleuvalserglupheryslnlyscysasbpnglnlas	1147
Oy	7545	TAGTGGACCTAAATATATATCTTGGATCCCTTAAGAACTTAATGCACTGATAAATATAT	7604
Db	1147	nlellelelglnashnrthylslysgln-ValArgsmleuasnlaqlueuasnlaas	1167
Oy	7605	AGGAAGAAGAAACCCCTTGAAAATAGAGAAAGAAATCGTTTTAAAGTAGAT	7657
Db	1167	erlysglnlylsalrg--Arglgylglnlysalglnleuenglnlaqlnvalg	1186
Oy	7658	----ATGAATGGAAATGTTACCAAAAATTCGAAGTTCATCAGAGAAA-----	7702
Db	1186	IngluleuasnAsnArgleuylslyvalglnAspqlueuHlsleuysThrIlelgln	1206
Oy	7703	-----AAGAGTATGTACTCCACAGAAAGAAC-----AAT	7736
Db	1206	luginmetThrHsArglysmetvalleuPheInlueIuserglLysPheylsglns	1226
Oy	7737	GTGCTTAAGCAATTTAGATGAAATTTGAAAATTCGAACACTTAAGATAGTAAATATCTCT	7796
Db	1226	eraln-glucInuphearg-Lys--LysmetglulysleuMetgluserlysValIleth	1244
Oy	7797	AAAAATGTTCTCGTGCAGACTGCAGAAATGAGAAATGACATATATATAAAAACTTCACCT	7856
Db	1244	rglnAsnAspIleSerlyIlelearg-----LeuAspPheVal--SerLeuIngl	1260
Oy	7857	AGAGAAACGGGCGCCATGCAAT-----CCAATATGTGTACTATGAATATATAGTTGCG	7910
Db	1260	nglnAsnSerArglaqlnlnlnAsnlnlaIylsleucysgluThr	1274
Oy	7911	TGATCTGGGTGCATCTAGTATGAGCAACAGATATGTTACGAATGTGCTTACTACCTCC	7970
Db	1275	AsnIlelysglnleuendlnuArg-----	1281
Oy	7971	CGTGAATATAAATATATATAGTTTTTGAAATACATATATGAAAAATGAGAAATAAAAA	8030
Db	1281	-----	1281
Oy	8031	TAAAGTAGAAATATATATACAAACAGATGTACAAACGTTTCGTTGCGTGAGATCTAA	8090
Db	1282	-----GlnleuInglInyArglnleuMetGlnInglYgl	1294
Oy	8091	TAGAAAAAGATATTTGGAAACCAATGACGTCGCAAGCAACCAAGAAATGCAAACTTTTGA	8150
Db	1294	nHsMetcglAlaAsnHsIlyTglnIlyscysglnlyleuGlnAspqlueuIlealagl	1314
Oy	8151	AAAAAGAAAGATGGATGATTTAGACGATATACATTAATACAAAGATTAAGTGTGCATTA	8210
Db	1314	nlyArgglValgluAsnleuylsgln-----Ly	1324
Oy	8211	GGACGATCCACCTGTGATGATATATATACCTCAACGCTTTCGATGAGATGACTGAATGCTC	8270
Db	1324	smetAspqlnqlnleuylsglnHs-----	1332
Oy	8271	TGAATATATATTTGAAACAGCTGTAGAGAAATGGAATTAATAAATCATCTGTGATCA	8330
Db	1333	-GlnHsIglInleuValleuIeuInglysglnlyserThrAlaIylas	1351
Oy	8331	CTGTAAAAACATCTGCACGATGCAAGATGATTAATGATGAATAAAGTGTACACACTGTAA	8390
Db	1351	pCys-----ThrPheIysProAspPhe-----GluMetThrVa	1362
Oy	8391	AACGAGATGCCAAGAAAT-----	8409
Db	1362	llysglnleuIlnHsIsergluleuSerSerArgasnThrGlnHsleuHsProth	1382
Oy	8410	-----AAAAATTTTCTTAAATGGAATCTCATCTGATATACATCAAAATTAATACAA	8465
Db	1382	rProArgSerProleuIeuHsIlyThrIlgln-----GluProGlnProleuIgl	1398
Oy	8466	AGATTTGATATACAAACCAATATATATACAAAATCTCTACTTATGATCATGTTCAAAATTT	8525

Dd	1398	uclnlystrpelnhihsrvalvalgluglnllepro-----lysgluvalgln--Ph	1415
Oy	8526	TGTCAAAAGTTCGAAACCTTTTAAAGTCAGTCTGTGAGACCTTTTGAATATCT	8585
Dd	1415	eglnproproglvalaerproleuulnlysgllysserserglnclytyrserglutyrph	1435
Oy	8586	TCATGAAACAAGTAAAGTGTGAATTATAATTATGAAATGATGGTCTCC----	8640
Dd	1435	esergethserthrcgluuglnllethrhphesrgluthrpsnprolllethragle	1455
Oy	8641	-----AATATCGAACATATGCTTTCGAAGAAC-----CCAAAAAGTTA	8681
Dd	1455	usergluilegluylsrleargspglnalleuanaenserserargproprovalargty	1475
Oy	8682	TAAAGAA---CGTTGACGTTGTACATA-----	8706
Dd	1475	rglnaspsnmalacysegluwtcgluuegluvallyslvalleuthrproleugluillealy	1495
Oy	8707	-----CGTTGTAGAAATCCATTT-----	8733
Dd	1495	asnllysglntyrasmethsrthrcgluvaltthrthrluylsnglulnlyslasnprowa	1515
Oy	8724	GGATATATTGTCTTACCGCATGCAAAACAAGTGGATGTAAAGAA-----TTTCAAC	8774
Dd	1515	lproserla---gluclutlrpmetleuglucylsargalaseargllygluyleuylsy	1534
Oy	8775	TTTTACCTTCGTCGCAAGACATGATTATGATTAATTCCTTGAAATGGAACGCATACCT	8834
Dd	1534	sglyasrpheneuylsyrgllyleucluproglutlthrrheglnasnpheasrglyasphl	1554
Oy	8835	TGTTCTTAATAGTCAGATGATAAC-----	8859
Dd	1554	salacyservalargpsarplurpheyphedngllyleuarghslsthrvalthral	1574
Oy	8860	-----AAAGGTATGATGATCTCCAAA-----	8883
Dd	1574	aargclnleuvalglunlalysleuueusrmetarglthrlleuglnleuargleugl	1594
Oy	8884	-----ACAGACATTTATGATACAAACCTAT	8909
Dd	1594	yleuylsrhvalglucluglvalglnlysrthrluauulysrpheluithrlysalathse	1614
Oy	8910	CACTCGATATATATGAAAAGGTATAAGAA-----	8943
Dd	1614	rllealeglyleuylsrleucluserthrysgllysrilesrphelaseerialaagl	1634
Oy	8944	-----ATTTTAAAAAAAACCTCTTCACTCTGCTTCAGTCAGACAA-----	8988
Dd	1634	uarglleleleasrlysmetvalalaleuallereleuugluallaglnalleathrgl	1654
Oy	8989	-----TTGTTAGCTCAAAATATTAATTCGGAAGACGTTG-----	9024
Dd	1654	yphelleleasrprolleesergllygnlthryrservalgluasparlavalleuylsgl	1674
Oy	9025	-----TCCTTGGCGAATGAAA-----TA	9044
Dd	1674	yvalvalasproclunherarglleargleuclunlalagluylsrhalaialaialgltty	1694
Oy	9045	TAGTTATGCACATTTATTCGATATTAATTAAGAAGACGTATGATGACACTTCATTATC	9104
Dd	1694	rserlytr-----	1697
Oy	9105	TGAAAAAATTAATAAATATTTGAACATCAATTAAGAACGCCGAATTCGTAAACATG	9164
Dd	1697	rserlysrthrluuservalphe-----Glnalleketgluasargmetleu-	1712
Oy	9165	GTGGGAAAAATATAGACGTAGATATGCGACGCTATGTTATGTGATATTAATTTATAC	9224
Dd	1713	-----Asprgrglnlysgllysrhslleuugluale-----Glnllealse	1727
Oy	9225	TTCAAAAGTAACTATTAATGATGAAGATGCGTCAATTACCAAAAGATGAAGAACTAATCA	9284
Dd	1727	rglgllyvalleasprvalargllyleargllyleargllyleargllyleargllylearglly	1747

```

: GENERAL INFORMATION:
: APPLICANT: Chisheti, Athar
: APPLICANT: Oh, Steven
: APPLICANT: Liu, David
: APPLICANT: Goel, Vikas
: APPLICANT: Li, Xuerong
: TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
: FILE REFERENCE: SI237/7019
: CURRENT APPLICATION NUMBER: US/10/087,464
: CURRENT FILING DATE: 2002-03-01
: PRIOR APPLICATION NUMBER: US 06/272,930
: PRIOR FILING DATE: 2001-03-02
: NUMBER OF SEQ ID NOS: 59
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 10
: LENGTH: 1639
: TYPE: PRT
: ORGANISM: Plasmodium falciparum
: US-10-087-464-10

```

Pred. No.:	2,01e-10	Length:	1639
Score:	289.00	Matches:	330
Percent Similarity:	33.65%	Conservative:	240
Best Local Similarity:	19.48%	Mismatches:	605
Query Match:	1.49%	Indels:	520
DB:	9	Gaps:	76

Db	198	AsnAspIrryCysGlnIleProPheAsnLeuLysIleArgAlaAsnGluLeu	-----ACAGCAGCAAAAATATTGGAAGATGTG	4515
OY	4516	CAGGAAAATATGT	AACTCAATGACCAATTGAAGTGTGATTACAGAA	4560
OY	217	LeuIysIysLeuValPheGlyTyrArgLysProLeuAspAsnIleLysAspAsnVal	-----	235
OY	4561	TGTATATAGAAATGCGACGACCTACGTTAAATATATGAAAAA	-----	4602
Db	236	-----GlyLysMetCysAspTyrIleLysLysAsnLysLysThrIleGluAsnIleAsn	-----	253
OY	4603	-----AAAAAGATGAGTCCCAAGATAAATATTATCAAGATGACCC	-----	4647
Db	254	GluLeuIleGluGluSerLysThrIleAspLysAsnLysAsnAlaThrLysGluGlu	-----	273
OY	4648	GACAAAAAAGATTGATGATGACACACATTGGTGTATGCTTACAGACTATATCTGGAACG	-----	4707
Db	274	GluLysLysLysLeuTyrGlnAlaGlnTyrAspLeuSerIle	-----	287
OY	4708	AATGCACACAGATTACTTGAAACAGAAATTTACTGCTAGTTGGTGATTAAGCCTCGAAGT	-----	4767
Db	288	-----TyrAsnLysGlnLeuGluGluAlaHisAsnLeu	-----	298
OY	4768	GCCTCTGTGTACAAAGAAATATACAAATGTATAGAAAACAAGGCTTACTATGATGCCGAC	-----	4827
Db	299	IleSerValLeuGlnLysArgIleAspThrLeuLysLys	-----	311
OY	4828	AAACATTTGGGGTGCACAAATTTATTGAAAAATGACGACAAATATATCAATTTTCAGT	-----	4887
Db	311	-----	-----	311
OY	4888	AAAGATTAAGTGCAAAAGATTAGTAAAGGAGGCAACACAGCTGCTATTAAATGGCAAAAC	-----	4947
Db	312	AsnGlnAsnIleLysGluLeuLeuAspLysIleAsn	-----GluIleLysAsn	327
OY	4948	AAAGTCTTATTAATACATAAATACTTGAAGAAATGACTGAAGATGCTCTTTTCCCTCT	-----	5007
Db	328	ProPheProAlaAsnSerGlyAsn	-----ThrProAsnThrLeu	340
OY	5008	CGTCGACTACTATATGTTTTTCATGCATTTGGATGGCAGCAATATATACAGATCCAGAAATTAAA	-----	5067
Db	341	-----LeuAspLysAsn	-----Lys	345

QY 5068 GATGAAATGGGTTGGCAAAAGATTGATGAGTGGCGCAACGGAAGGTACATTTG 5127
DB 346 LysIleGluGluHISgluLysGluIleLysGluIleAlaLysThrIleLysPheAsnIle 365
QY 5128 GGT-----CAATACTACAAAGAAAAAGAAAAAGAG 5160
DB 366 AspSerLeuPheThrAspProLeuGluLeuGluTyrTyrLeuAlaGluLysAsnLysAsn 385
QY 5161 -----AAATTAATAAGCGCATGGCGCAAAATATCTATGAGTGGCGG 5205
DB 386 IleAspIleSerAlaLysValGluThrLysGluSerThrGluProAsn---GluTyrPro 404
QY 5206 CCTTGATGCTATGCAAAATATAGTTTATGATTAAGATATAATCTAGATATTGAT 5265
DB 405 AsnGluValThrTyrProLeuSerTyrAsnAspIleAsnAlaLysGluLeuAsn 424
QY 5266 AATTGGGAAGAT-----GAAAAACAAGACCGAGAAATTTGAAAGAAATATT 5316
DB 425 SerPheGluAspLeuIleAsnProPheAspTyrThrLysGluProSerLysAsnIleTyr 444
QY 5317 AACAAATAGCAACATCACTGTGCAAGAGATGATAGTACAGAAATCCCGTAGT 5376
DB 445 ThrAspAsn----- 447
QY 5377 ACTGCGCGAAATTTTCTGCAACGAAATAGATGTGTGCAACGCAATGATATGC 5436
DB 448 ---GluArgLysLysPheIleAsnGluIleLysGluLysIle-----LysIleGlu 463
QY 5437 GGGTCAACAAACGTGTGATGATGAAATAGTGAATAGTGCAGAAAGTGTAGTAAAT 5496
DB 464 LysLysLysIleGluSerLysLysSerTyrGluAspArgSerLysSerLeuAsnArg 483
QY 5497 CTAAAAAATGCTGCTCTACCTTCAGATGATATATGCGGAAAAATCCCGAT 5556
DB 484 IleThrLysGluTyrGluLysLysLeuAsnGluIleTyrAspSerLysPheAsnAsn 503
QY 5557 -----GAAGTACTGCTATCAGTTTCTT----- 5580
DB 504 IleAspLeuThrAsnPheGluLysMetMetGlyLysArgTyrSerTyrLysValGluLys 523
QY 5581 -----CGATGCTTCCCGAATGGCGGAAGATTTTGCAGAAATATAAGAAAG 5628
DB 524 LeuThrHisHisAsnThrPheIleAsnTyr-----GluAsnSerLysHisAsnLeuGlu 541
QY 5629 GAATGCGAATAATGCTAGCGCGCTGATATGATTAATCTGCTGATATGATGAAGATATA 5688
DB 542 LysIleThrLysAlaLeuLysTyrMetGluAspTyrSerLeuArgAsnIleValIleGlu 561
QY 5689 AGAAGAAATGTACAGATCGGTACACAAATTAATAAATTTATTAAGAGTGGAAACCA 5748
DB 562 LysGlu-----LeuLysTyrTyrLysAsnLeuIleSerLysIleGluAsn 576
QY 5749 CAGTATGAAAAACAATCAAAAAATATGCTGAGAAATTAAGACAAATATATCCGAGCAT 5808
DB 577 GluIleGluThrLeuValIleLysAsnIleLysLysAspGluGluLeuPhe---GluLys 595
QY 5809 CCTGTGCAAAAGATGCAAGAGACGCTCCGAAATATTAGACAAACATTAATAAATTT 5868
DB 596 LysIleThrLysAsp----- 600
QY 5869 TGTGAAAAATAAAGTGGAGAT-----TGTGAATATATAGTATGAAGATGTGCCACA 5922
DB 601 ---GluAsnLysProAspIleLysIleLeuGluValSerAspIleValLysValGluVal 619
QY 5923 CAGCGATTAACGATGATAGTCAAAATATGATCCCGCATCATATAGACATGAACCAAA 5982
DB 620 GlnLysValLeuLeuMetAsnLysIleAspGluLeuLysLysThrGlnLeuIleLeuLys 639
QY 5983 GAAGTTGAAGAAAGTGTATTTGCAAGTGCACAGAGTCCACAGCTGTACGAAAGGAA 6042
DB 640 AsnValGluLeuLysHisAsnIleHisValProAsnSerTyrLysGlnIleLysAsnLysGln 659

QY 6043 ACACGCTCACACGGGTATCACTGATATCAAAAGCAGCGCATCGAAAAAGAA----- 6096
DB 660 GluProTyrTyrLeuIleValLeu-----LysLysGluIleAsp 672
QY 6097 ---GCGAAACACGCGCGCTACAAACACCGCAAAAAAGTGAATAATCAACACGAA 6153
DB 673 LysLeuLysValPheMetPro-----LysValGluSerLeuIleAsnGlu 687
QY 6154 ATGCGGACCAAAACAGAACCCGACAGACGACCAACAAACGAAACGAAACGATCAACA 6213
DB 688 GlnLysLysAsnIleLysThrGluGluGlnSerAspAsnSerGluProSerThrGluGly 707
QY 6214 GCAACACACACAGATCTGACGTGGCACATGTAAAGCCATCTTCGATTAATACCA 6273
DB 708 GluIleThrGlyGlnIle-----ThrThrLysPro 717
QY 6274 GATGACGAGGGTGA-----ATAGAGGTTGTAATCCAAAAACGTATGACCAATATCCT 6327
DB 718 GlyGlnGlnAlaGlySerAlaLeuGluGluLysAspSerValGlnAlaGln----- 735
QY 6328 AATGCGGTTGATGTAGTAAAGTAAAGAAATGAAGAAATGGCATATGCTGCT----- 6384
DB 736 -----GluGlnLysGlnAlaGlnProProValProValProVal 748
QY 6385 CTTAGAGAAAAAAT 6444
DB 749 ProGluAlaLysAlaGlnValProThrProProAlaProValAsnAsnLysThrGluAsn 768
QY 6445 AACCGTCAACATGATATAAAGACGCTTTTATTAATGTGCAACATACAACTCAATT 6504
DB 769 ValSerLysLeuAspTyrLeuGluLysLeuTyrGlu-----PheLeuAsnThrSerTyr 786
QY 6505 TTGTGTTAAAT 6558
DB 787 IleCysHisLysTyrIleLeuValSerHisSerThrMetAsnGluLysIleLeuLysGln 806
QY 6559 GGAACAATTCACAGATGATTAAGAAATATGATATATATATATATATATATATATAT 6618
DB 807 TyrLysIleThrLysGluGluGluSerLysLeu-----SerSerCysAspProLeuAsp 824
QY 6619 ATGTTTGTGGAACATATTTCTATGATATAAATAATATA-----ACT 6663
DB 825 LeuLeuPhe-----AsnIleGlnAsnAsnIleProValMetLysSerMetPheAspSer 842
QY 6664 GTAACAATAGTGTACACACATCTCAATGAAATATAAAGAAAGAA----- 6711
DB 843 LeuAsnAsnSerLeuSerGlnLeuPheMetGluIleTyrGluLysLysMetValCysAsn 862
QY 6712 -----CAGGATAAAAAAAGAT-----GAGCAATTAAGTAAATA 6747
DB 863 LeuTyrLysLeuLysAspAsnAspLysIleLysAsnLeuLeuGluGluAlaLysVal 882
QY 6748 TTTTGGGAAAAAATATAAATTTATTTGGAAGAAATATATATATATATATATATATAT 6807
DB 883 -----SerThrSerValLysThrLeuSerSerSerMetLeuProLeuSer----- 898
QY 6808 CTCACAGACGAAACGAAACGAAACGAAATTAAGATTAATACAGATGATGATGACG 6867
DB 899 LeuThrProGln---AspLysProGluValSerAlaAsnAspAspThrSerHisSerThr 917
QY 6868 AAATGACGCTTCCCTTGAAGATTT----- 6894
DB 918 AsnLeuAsnAsnSerLeuLysLeuPheGluAsnIleLeuSerLeuGluLysAsnLysAsn 937
QY 6895 -----GTAAAAAGCCCAATTTTGTGATGATGCTTACAGAA 6930
DB 938 IleTyrGlnGluLeuIleGlyGlnLysSerSerGluAsnPheTyrGluLysIleLeuLys 957
QY 6931 TGGGCGAAGAAATTTGTAATAGAGAGAAAGCAACACTTTTAAATATGAGCGGCGCTGT 6990
DB 958 AspSerAspThrPheTyrAsnGlnLysSerPheThrAsnPheValLysSerLysAlaAsp----- 976
QY 6991 AAGCAATATGAGTGTATGATGATTAATGACGCTTAAGACACAGAAATGTGACAGGCGTGT 7050

Mon Apr 28 13:47:29 2003

us-10-087-013-1.n2p.rappb

Page 41

```
QY 8763 GGATTACAACTTTACCTTCCTGCTCGAAGATGATTAT 8802
      |||||
Db 1615 -----Ilephcysersersersanphe 1622
```

Search completed: April 28, 2003, 13:38:05
Job time : 717 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus.n2p model

Run on: April 28, 2003, 11:02:51 ; Search time 96.5 Seconds
(without alignments)
6480.970 Million cell updates/sec

Title: US-10-087-013-1
Perfect score: 19435
Sequence: 1 atggggctctctgcgaata.....atatatgatgaagatg 10628

Scoring table: BLOSUM62
Xgapop 10.0, Ygapext 0.5
Fgapop 10.0, Fgapext 0.5
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame.n2p.model -DEV=xlp
-Q/cgn2_1/USPTO.spool/US10087013/runat.28042003.102846.20293/app.query.fasta_1.10823
-DB=Issued_Patents_AA -QWMT=fastan -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=BLOSUM62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10087013.ecgn.1.159.etrnat.28042003.102846.20293 -NCPU=6 -ICPU=3
-NO_XUPXY -NO_MAP -LARGEUSER -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/6C.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	3501	18.0	3060	2	US-08-487-826B-14
2	2972.5	15.3	2710	2	US-08-568-459A-12
3	2972.5	15.3	2710	2	US-08-487-826B-12
4	2972.5	15.3	2710	4	US-09-210-288-12
5	1755	9.0	2182	2	US-08-487-826B-16
6	1032.5	5.3	700	2	US-08-568-459A-10
7	1032.5	5.3	700	2	US-08-487-826B-10
8	1032.5	5.3	700	4	US-09-210-288-10
9	649.5	3.3	921	2	US-08-568-459A-8
10	649.5	3.3	921	2	US-08-487-826B-8
11	649.5	3.3	921	4	US-09-210-288-8
12	618	3.2	1435	2	US-08-568-459A-4

13	618	3.2	1435	2	US-08-487-826B-4	Sequence 4, Appl1
14	618	3.2	1435	2	US-09-210-288-4	Sequence 4, Appl1
15	431	2.2	749	2	US-08-568-459A-6	Sequence 6, Appl1
16	431	2.2	749	2	US-08-487-826B-6	Sequence 6, Appl1
17	431	2.2	749	4	US-09-210-288-6	Sequence 6, Appl1
18	418	2.1	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
19	384.5	2.0	3135	1	US-08-323-170B-2	Sequence 2, Appl1
20	384.5	2.0	3135	4	US-08-954-441-2	Sequence 2, Appl1
21	383	2.0	3248	5	US-08-353-700-1	Sequence 1, Appl1
22	383	2.0	3248	5	PCT-US95-16216-1	Sequence 1, Appl1
23	365	1.9	411	2	US-08-568-459A-20	Sequence 20, Appl1
24	365	1.9	411	2	US-08-487-826B-32	Sequence 32, Appl1
25	365	1.9	411	4	US-09-210-288-20	Sequence 20, Appl1
26	349.5	1.8	3596	4	US-09-134-001C-5080	Sequence 5080, Ap
27	343.5	1.8	1115	2	US-08-568-459A-2	Sequence 2, Appl1
28	343.5	1.8	1115	4	US-08-487-826B-2	Sequence 2, Appl1
29	343.5	1.8	1115	4	US-09-210-288-2	Sequence 2, Appl1
30	343.5	1.8	1115	6	5198347-6	Sequence 2, Appl1
31	286	1.5	2482	1	US-08-328-254-6	Sequence 6, Appl1
32	287.5	1.5	3418	2	US-08-603-753D-4	Sequence 4, Appl1
33	287.5	1.5	3418	4	US-09-099-753-4	Sequence 4, Appl1
34	287.5	1.5	3418	4	US-08-986-106-4	Sequence 4, Appl1
35	283	1.5	3418	2	US-08-639-501-2	Sequence 2, Appl1
36	283	1.5	3418	3	US-09-044-946-2	Sequence 2, Appl1
37	283	1.5	3418	3	US-09-044-908-2	Sequence 2, Appl1
38	280	1.4	3418	3	US-08-755-587-44	Sequence 44, Appl1
39	279	1.4	362	2	US-08-568-459A-18	Sequence 18, Appl1
40	279	1.4	362	2	US-08-487-826B-30	Sequence 30, Appl1
41	279	1.4	362	4	US-09-210-288-18	Sequence 18, Appl1
42	276.5	1.4	2285	4	US-09-308-375-2	Sequence 2, Appl1
43	276	1.4	411	2	US-08-568-459A-19	Sequence 19, Appl1
44	276	1.4	411	2	US-08-487-826B-31	Sequence 31, Appl1
45	276	1.4	411	4	US-09-210-288-19	Sequence 19, Appl1

ALIGNMENTS

RESULT 1
US-08-487-826B-14
Sequence 14, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sam, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent-In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487, 826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH21.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3060 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-487-826B-14

Alignment Scores:

Pred. No.:	7,096-258	Length:	3060
Score:	3501.00	Matches:	1057
Percent Similarity:	38.65%	Conservative:	470
Best Local Similarity:	26.75%	Mismatches:	934
Query Match:	18.00%	Indels:	1490
		Gaps:	147

US-10-087-013-1 (1-10628) x US-08-487-826B-14 (1-3060)

```

25 ATTTAATAATGGGAATGCACATCATTTAGAGGAGCTTAAGCCCTATTATA 84
   ::::::::::::::::::::
4  LeuAlaLysMetGlyProLysGluAlaIaGlyLysAspIleGluLys----- 20
QY  AAAGAAAGTCACAAAGCTGCAAGAAATGTTTGAAGCTTATGCCAAATATAGACAT 144
   ::::::::::::::::::::
21 -----GluSerAlaLysHisMetLysPheAspArgIleGlyLysAspValTyrAsp 36
QY  145 CCATCAAAA---TATGCAAAAAGACATGTGATTCGTGAAAGGGGATTTCAGCAAGA 201
   |||
37 LysValLysIleGluLysAlaLysGluArgLysGlyLeuGlnGlyArgLeuSerGluAla 56
QY  202 GAATTCGGTGGTGGCTCCTACGCCAGTAATTAACATTAATTTATTATTCATATCCA 261
   ::::::::::::::::::::
57 LysPheGluLysAsnGluSerAspProGlnThrProGluAsp-----Pro 71
QY  262 TGTAAATTTAGATCATAGACACATTAATTTACGGTATGATGATGTAATTTGAGACAT 321
   |||
72 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 88
QY  322 CCTTGCATGCTAGACAAACCAAGCTTTGATGAAGATGAAGAAATTCGATGT---GGA 378
   |||
89 ProCysAlaAspSerArgSerAspValArgPheSerAspGluTyrGlyGlnCysThrHis 108
QY  379 AATTAATACGTAAATTAATAAGAAATAATGCTATATACCTGTGCGCCACTGAGAGA 438
   |||
109 AsnArgIleLysAspSerGlnGlnLysAsnLysGlyAlaCysAlaProTyrArgArg 128
439 CGACATATGTGATTAATAAACTTGAACGCTTAATGATTAATAATACCAAAATTTTCAT 498
   |||
129 LeuHisValCysAspGlnAsnLeuGlnIleGluProIleLysIleThrAsnThrHis 148
QY  499 GATTTATTTGGGAATGTACTAGTACCAAAATTCGAAGTGAATCAATGTTAATATAT 558
   ::::::::::::::::::::
149 AsnLeuLeuValAspValCysMetAlaAlaLysPheGluGlnSerIleThrGluAsp 168
QY  559 CATCCA---CATAAAGAACT-----TCAGACGCTTGACTGCTCTTGA 600
   ::::::::::::::::::::
169 TyrProLysTyrGlnAlaIleThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 188
QY  601 CGAAGTTTGGAGATATGCTGATATTTAGAGCAATAGATATGTTAAACCAATATGTC 660
   |||
189 ArgSerPheAlaAspIleGlyAspIleValArgLysArgLysLeuTyrLeuGluLysAsnPro 208
QY  661 CATGAC-----AAAGTAGAAACGGGCTCCGAGAGGCTTTCAAGAAATA 705
   ::::::::::::::::::::
209 GlnGluIleLysGlnArgGlnGlnLeuGlnAsnLeuLysThrIlePheGlyLysIle 228
QY  706 CATGAT-----GGAATGGAAGATGAAGTAAATAATGATTAACAATCTGATGATCT 756
   ::::::::::::::::::::
229 TyrIleLysLeuAsnGlyAlaGluAlaArgTyrGlyAsnAsp-----ProGlu----- 244
QY  757 GGAATTTATTATAATTAAGGAAGCATGTGCTGAATCTGAATGAATTAAGTATGGGA 816

```

```

Db 245 -----PheHeLysLeuArgGluAspTyrPThrAlaAsnArgLysThrValTyrLys 262
QY 817 GCTATAACATGATGATGCATCATATTAATCTGGATATTTATGCAATCAAGAAATATCA 876
   |||
Db 263 AlaIleThrCysAsnAla---TrpGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 280
QY 877 CCATTATTTTCAAAATCCGCGCCATTAACAAAGCAAGAGCTTCACCAATTTAAT 936
   ::::::::::::::::::::
281 GlyLysArgThrLysGlyTyrCysArgCysAsnAspArgLysValProThrTyrPheAsp 300
QY 937 TATGTCCTCATATATTTAGCTGCTGCTGCGACGATGGGAGAGACTTTGCCGAAAGA 996
   |||
Db 301 TyrValProGlnTyrLeuArgTyrPheGluGlnLysThrPalGluAspPheCysArgLys 320
QY 997 AATTAATTAATGAAGAAAGCTCAAGACCTGCTGCT-----AATGACAAAGAA 1044
   |||
321 AsnLysLysIleLysAspValLysArgLysAsnGlyLysAspLysGluLysAsp 340
QY 1045 CGCTTATATTTGATCATATATGACATGATTTGACGACAACTAATTCGAAAGCTATT 1104
   |||
341 Arg---TyrCysSerArgAsnGlyTyrAspCysGluLysThrLysArgAlaIleGlyLys 359
QY 1105 TTGCATTTGCAATTAATAGCTAGCTGCTGCTGCACTAATTCGAAAGCTTTTGAAGTTGG 1164
   |||
360 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTyr 379
QY 1165 TTAGGAATTCACAAAGACATTTAAACAAAGAAATAATGTAAGAAATATCA 1224
   |||
380 IleAsnAsnGlnLysGlnGlnPheAspLysGlnLysLysLysTyrAspGluGluIleLys 399
QY 1225 TCATAT-----TTATCGAACGATTAACAATTTGTCAT 1257
   |||
400 LysTyrGluAsnGlyAlaSerGlyLysArgGlnLysArgAspAlaGlyThrThr 419
QY 1258 AATTAATTAAGTAA---TATTATAACAATTTATGAAAACTTAAAGAAAGCAATAT 1314
   |||
420 ThrThrAsnTyrAspGlyTyrGlyLysLysPheTyrAspLeuAsnLysSerGlnTyr 439
QY 1315 GCACATATGACACTTTTAAATTTACTTAATGAAAGAAATGATTTGT----- 1362
   |||
440 ArgThrValAspLysPheLeuGlnLysLeuSerAsnGlnIleCysThrLysValLys 459
QY 1363 -----AAAGAGAGATTCACGAGAAAGATATTAATTTACTTACACGCTGT----- 1410
   ::::::::::::::::::::
460 AspGluGlnGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 479
QY 1411 -----GATGCAAAAGGATATTTATCGTTCAAGATATTCGCAAGTGT 1455
   ::::::::::::::::::::
480 SerGlyThrAsnValGluSerGlnGlyThrPheTyrArgSerLysTyrCysGlnProCys 499
QY 1456 CCCGACTGCGGGGTCAAAATGATGATTAATAATACACACAC---AATACATTAATGAT 1512
   |||
500 ProTyrCysGlyValLys-----LysValAsnAsnGlyCylSerSerAsnGln 515
QY 1513 CGTGAACGTGAATATAT-----GAAGACTATTAACCTCCATGGGGT 1554
   |||
516 TrpGluGlnLysAsnAsnGlnLysCysLysSerGlyLysLeuTyrGluProLysProAsp 535
QY 1555 GTGACGCTAGTAATATTCACCTGCTTATAGCTGATGAACAAAGCTGATATACACA 1614
   ::::::::::::::::::::
536 LysGlnGlyThrThrIleThrIleLeuLysSerGlyLysGlnHisAspAspIleGluGlu 555
QY 1615 AATTAAGAAATTTTGTATAC-----AGCTCA 1641
   |||
556 LysLeuAsnLysPheCysAspGluLysAsnGlyAspThrIleAsnSerGlyLysIleSer 575
QY 1642 ACTAATTAACAAAGATTAATAATAT-----CAAAAATGGAATGTC 1680
   |||
576 ThrGlyGlySerGlyGlyLysAsnSerGlyArgGlnGluLeuTyrGlnIleThrLysCys 595
QY 1681 TATTATTAAGGATGAATAATTAATGATGATTAACGGAACAAATTAATGATCAATTAAT 1740
   |||

```

Db	596	---	yytlytsglysluAspValVallyslValGlYhNlsAspRluAspRluclunspRlyr	614
Oy	1741	GATAACTCTAAG	-----ATAATA-----	1758
Db	615	gluAsnVallyslAsnlaGlYglYleuCySlleleuYlAsnGlIndlyslAsnlyslu	634	
Oy	1759	-----	-----TCATTCATTAATTTT	1773
Db	635	GluclyclYAsnThrSerGlYslGuPProAspRlIlleGlIndlyslThrPheAsnProPhe	654	
Oy	1774	TTTGAATTAATGCGTTACAACTTTATTTAATTAAGGAACTACATTAAGTGCATCAACAATCTTAA	1833	
Db	655	PheYrlyrlyrtrpValAlaHlsMetleuYlAspSerIlleHlsTrpYlslYslslucln	674	
Oy	1834	ACTTGTATA--AATAATACACACCAGCATGT--ATTGATGAATGTACAGAAATGC	1887	
Db	675	ArgCySlleuGlInAsnGlYlAsnAArglIleYslCyGlYlAsnIndlyCyAsnHlsAsn	694	
Oy	1888	TTATGTTTGTGACAGATCGGCTTAAACAAAAAGAAAGATGCAATATGATTAAGAAACTG	1947	
Db	695	glucysPheYlAsnArgTrpIletrHnGlIndlyslAspRlItrpGlYslleValGlIndHls	714	
Oy	1948	TTTCAAAAAAAGATTAATACAG	1971	
Db	715	Phe---lyslThrGlInAsnIlleYslGlYArgGlYlSerAspAsnThrAlaGluleuIle	733	
Oy	1972	-----CAATCGTATTTATAGTAATATTAAATCTTTTGAAGTTATTTT	2016	
Db	734	ProPheAspHlsAspYrValleuGlIndYrAsnIndleuGlIndlupheleuYslYAsp	753	
Oy	2017	-----TTTAAAGTTATGATTAACCTTGACAAAGATGACCAAAATGC	2058	
Db	754	SerGlIndspAlasSerGlIndYlYslSerGlIndAsnSerleuAspAlaGlIndlula	771	
Oy	2059	AAAGAACTTATGCAATATATTAATAAGAAAAAAGAGTTTCCATTTGCAATTTGAT	2118	
Db	772	-----GlIndleuYlslHlsleuArgGlIndleuSerIndspAsnHn	787	
Oy	2119	AGGCAC	-----TATTAGAAATGCATTAAGACTCTGTTA	2154
Db	788	GlIndlAserValGlYglYlValThrGlIndlYslAsnIlleMetAspYslleu	807	
Oy	2155	GATCACTTAAAAAGAACTGCCACGATGTG	2184	
Db	808	AsnYrGlYlYslAspRlIulAspLeuCySlleuGlIndleHlsGlIndspGlIndlun	827	
Oy	2185	---AAAGCATATATTAACAAGAGATGTGAACAATCCCATTAATGCACAAACAAACCG	2241	
Db	828	lyslGlIndYslglYAspRlYAsnGlYslleGlIndlYglIndspHeArGTrYAsnPro	847	
Oy	2242	TGTGTTAACTCTGTGAGGCACAGCA--CCCACTTAAATATTAAGAAATATACACA	2298	
Db	848	CysSerGlYglIndserGlYAsnYlYslArgYrProValleuAlAsnYslValAlaTrYglInd	867	
Oy	2299	TACTTTAAAGAGACTGCATACAGAGAAACACAGAAATCGTGCTTCATTAATTTGAAGCA	2358	
Db	868	MetHlsHlsYslAlaTrYslThrGlIndleuAlaSerArGlAlYArgSerAlaLeuArgYl	887	
Oy	2359	AAGGACACAGAAAGTATATTAACAAAGCGGGCGGTAGAGAAAGAGACTCAAGACAAATTA	2418	
Db	888	AspIleSerleuAlaGlIndPheYlYslAsnGlYlYArgAsnGlYSerThrleuYslGlIndle	907	
Oy	2419	TGTGAATTAATGATTAACAACTTAAATCGTAATCTTGTTTTCAAATGACCATGTGAT	2478	
Db	908	CysYlslleAsnGlIndYrSerAsnAspSerArgGlYlAsnSerGlYlYrProCySlThr	927	
Oy	2479	GGCAAGGCACAGGTGATGATTAACAACAAGATTTGTCTGAGAACTGAATGCAAGATG	2538	
Db	928	GlYlYsl---AspRlYAspHlsGlYlYValArgMetArglIndYlThrGlIndYrPserAsn	946	
Oy	2539	GATCGGAAACACATGCGCTAAAGATCAACGAAGATGTTATATGCTCTCTGAAGACAGACAT	2598	
Db	947	IlleGlIndlYslYslIndThrSerYrlylYslAsnValPheleuProTokrArgRglYslH	966	

QY	2599	ATAGTACATCCATTGTTGGAAACATTTCACAAAGGATGATCACCACCTTAAGCGTAATATT	2658
Db	967	MetcysinSerAsnLeuGlnAsnLeuAspVal	980
QY	2659	GTT---GATGATTAGTAAATTAATTCCTTTTGGGGATGTTCTTATCAGCAAAATAT	2715
Db	981	ThrlvsAsnAspLysAlaSerHisLeuLeuGlnLysPvalGlnLeuValLysThr	1000
QY	2716	GAGCAAAACAGATTAATATCCAAAGCTATTAAGAAAAGAAATATACCTTAAGAGGCCCAAGAA	2775
Db	1001	AspAlaAlaGlnIleIleLysAsArgLysAspGlnAsnAsnIle	1016
QY	2776	GTAATGCACCA-----AAACACACACAACTATCTGCGAGCTATACGTTACGTTACAGT	2826
Db	1017	LeuthrAspProIleGlnGlnLysAspGlnGlnLysAlaMetCysArgAlaValLArgTrSer	1036
QY	2827	TTTCAGATATAGTGTATATATATTCGAGAGAAGATCTCTGGAAAGAAACGTCATGC	2886
Db	1037	PheAlaAspLeuGlnLysAspIleIleArgGlyArgAspMetTrpAspGlnLysSerSer	1056
QY	2887	GTAAGCTGCAAGGACATTTGGAAACGTTTGGTAATATCATATAGCACTCATAAAGC	2946
Db	1057	ThrAspMetGluThrArgLeuIleThrValPheLysAsnIleLysGluLysHisAspLys	1076
QY	2947	---AAGAGAAATGATTAATATTAATGATGAT-----GCCCCAAATATTTAAATATG	2994
Db	1077	IleLysAspAsnProLysTrpGlnLysAspGluSerLysLysProAlaLArgLysLeu	1096
QY	2995	AGGGAATTTGGTGGGAAGCTATATAGAGCCAAAGTGTGGAAAGCCATCAATGTGATATA	3054
Db	1097	ArgAlaAspTrpTrpGlnLysAlaAsnAlaGlnHisGlnValTrpArgAlaMetLysCys	1114
QY	3055	AAATATTTGAAGGATTAATATCGGGACACCAATCAACAAAGTAGTATGCGGATATAGT	3114
Db	1115	-----AlaThrLysGlyIleLeuLys-----	1121
QY	3115	GATCATACACCATTTGATGATTATATCCACAAAAATTAAGATGATGACGAGATGGCA	3174
Db	1122	ProGlnMetProValAspAspLysIleProGlnArgLeuArgTrpMetThrGluTrpAla	1141
QY	3175	GAATGCTCAGCAGGCGGCAAAAGAGATGATGATGATGAAAGAGAGAGCTAGAG	3234
Db	1142	GluTrpTrpCysLysAlaGlnSerGlnGluTrpLysPheLysLysLysLecLysAlaAsp	1161
QY	3235	TGTAAAGATTAAGATTAATGCTCAAGGCTGTACGAAAGAGAGTGTACAGGTGTACGAG	3294
Db	1162	CysMetSerLysGlnAspGlyLys---CysThr---GlnGlyAspValAspLysGlnLys	1179
QY	3295	TGCCACAGAACTTGTATGATATATAAGATTAATAGATTTATGCAAGACAAATGAT	3354
Db	1180	CysLysAlaAlaCysAspLysTrpLysGlnGlnIleGlnLysTrpAsnGlnLeuIleTrpArg	1199
QY	3355	ATAATATCAGATTAATATCAAAAGATTCATGAGCAACAAATGCTGTGTGTAATAGT	3414
Db	1200	LysIleSerAspLysTrpAsnLeuLeuTrpLeuGlnAlaLysThrThrSerThrAsnPro	1219
QY	3415	GGTATTAAGAGCTTCCAGTACTGCCAAAAATCATATAGAC-----AGAAATGTAT	3465
Db	1220	Gly-----ArgThrValLeuGlnLysAspAspAspProAspTrpGlnMetMet	1235
QY	3466	GAATTTTGTGCGAATTAATCCAAACAAATATGCGGCAAA-----AGTATATAA	3513
Db	1236	AspPheLeuThrProIleHisLysAlaSerIleAlaAlaArgValLeuValLysArgAla	1255
QY	3514	AGTGCTACTAGTGAATGAAGAGCTGTCATTTGTACTATACACACAGATGAATAATGTGGA	3573
Db	1256	AlaGlySerProThrGlnIleAlaAlaAlaAlaProIleThrProGlySerThrAlaAla	1275
QY	3574	GCATATCTCCATGATACAGAAATTTTGATGATGTTCAGTCAACAAATGAGTTTGTGAT	3633
Db	1276	GlyLysIleHisGlnGlnIleGlyLysGlyCysGlnGlnGlnIleThrClnPheCysGlu	1295

Db	1644	AspProclutrlYrlysglyTrglutrlYrLysasp-	165
QY	4723	TTGAAACAGAAATTTACTGCTACTGTGTGGATTAACCTCGAAGTGCCTCTGTGTACAA	476
Db	1655	-----GlyValGlnPro-----IleGln	166
QY	4783	AGAAATATACATGTTGTAGAAAAACAGSGTTACTACTGATGCCGACAAACATTGGGTGC	488
Db	1661	GlysnclutrlYrLeuclnlyns-----Cysaspasnlnlyns--CysSerCys	167
QY	4843	ACAAATTTATGTAAAAATGACGACAAATATACTAACATTTGAGTAAGATATAGTGCAA	490
Db	1676	-----	167
QY	4903	GGATAGTAAGAGGCAACACACAGTGCATTTAAGTGGCAAAACAAAGTCTAATPAAC	496
Db	1676	-----	167
QY	4963	TACAAATTACTGAAAGAAATTGACTGAGATGTGTTTTCTCTCGTGCACACTATATA	502
Db	1676	-----	167
QY	5023	TGTTTCATGCATGTGATGGCAATTATACAGATCCAGAAAGTTAAAGATGAAATGGGTG	508
Db	1677	-----MetaspelYasnValleuSerValserProlysglnlys-----	168
QY	5083	CGAAAAAGATGTATGGAAGTGGCGCACGAGAGGTCATTTGGGTCAATCTACAAA	514
Db	1690	-----PropieglyLysTyr-----	169
QY	5143	GAAAAAAGAAAAAGAGAAAAATTAATAAAGCTGGATGGCGACAAATATTTCTTAGAGTC	520
Db	1695	-----AlaHisLysTyrPro-----	169
QY	5203	CCGCTTGTAGTGTATGAAATATAGTTTTATGATTTAAGAGATATAATTTAGTATT	526
Db	1699	-----	169
QY	5263	GATAAITTTGGAAAGATGAAAAACAAAAGACCGAGGAAAAATTTGAAGAAATATTTAACAA	532
Db	1699	-----	169
QY	5323	AATGAAATCAGTGTGGCAAGGAAGTATAGTACTACAGAAATCCCGGTACTGCG	538
Db	1699	-----	169
QY	5383	CGAAAAATTTTCTGAAACGAAAAATTAAGAAATGTGTGTGAAACCAATGATATCGGGTAC	544
Db	1699	-----	169
QY	5443	AAACGTGTAGGATGATGAAATATAGTGAATATAGTGCAAAGAACTATGAACATCTAATA	550
Db	1699	-----	169
QY	5503	AAATGTCTCTGTACTTCCATGATGATTAATCTATAGGGAATAATCGCATGAAGT	556
Db	1699	-----	169
QY	5563	ACTGCGTATCAGTTTCTTCATAGTGTTCGCGAATGGGGTGAAGATTTTTCGAAACATATA	562
Db	1699	-----	169
QY	5623	GAAGAAGAAATTTGAGAAATTTGTAGGGCGCTGAATGATTAATACTTGTGCTGATATGAA	568
Db	1699	-----	169
QY	5683	GATAAAGAAAGAAATTTACAGATGCGGTGTACACAAATTAATAAATTTATTAGTGAATGG	574
Db	1700	-----GlnLysCys-----	170
QY	5743	AAACCAAGTATGAAAAACAAATCAAAAAATATGTGGAATTAAGACAAATATATATTC	580

Db 1702 ----- 1702
QY 5803 GAGCATCTGTGGCAAAAGATGACAGACGCTCGCAATATTATTAGACAAACAATTAAAA 5862
Db 1702 ----- 1702
QY 5863 AAAATTGTGAAAAATAAAAGTGAGATTGTGAATATAGTATGAAAGATGTGTCCACA 5922
Db 1703 ----- 1704
QY 5923 CACGATTAACGTATGTATAGTCAAAATATGCCCGCATATTAGACGATGACCAAAA 5982
Db 1704 ----- 1704
QY 5983 GAAGTTGAAGAAAGTATATTGTCAAGTCCACAGAGTCCACAGCTGTACGAGAGAA 6042
Db 1705 ----- 1715
QY 6043 ACACGCTCACACGGGTATCACTGATATCAAAAAGCAGGCGCATGAAAAAGAGCGAAA 6102
Db 1715 ----- 1715
QY 6103 ACAGCGCGCTACAAAACAGCGGAAAAAGTGAAAAATCTAACACAGAAATGCGAGCA 6162
Db 1716 ----- 1723
QY 6163 CAACAACGAACCCGACGACGACACAACAACAGAAACATCAACAGCAGCAACA 6222
Db 1724 ----- 1728
QY 6223 ACAGAAATCGAGCTGGCGCAATGTAAAGCCATTCTTCGAAATACCAACGATACGAG 6282
Db 1729 ValThrValAspValCysSerIleValIysThrLeuphels ----- AspThrAsn 1745
QY 6283 GGTGAATAGAGGTTGTATCCAAAACGATGACAAATAT ----- CCTAAATGGGGT 6336
Db 1746 AsnPhSerAspAlaCysGlyLeuLys ----- TyrGlyLysThrAlaProSerSerThrLys 1764
QY 6337 TGTATT ----- GTAGTAACTCTAAAGAAAT 6363
Db 1765 CysIleProSerAspThrLysSerGlyAlaIleThrThrcIlysserGlySerAsp 1784
QY 6364 GAAATATGCATATGATGCTCCTCTAGAGAAAAATTTATGTAATAATATATACATAT 6423
Db 1785 SerGlySerIleCysIleProProArgArgArgLeuThrValGlyLysLeuGln --- 1803
QY 6424 TTAATATTAAGAACT ----- GAAATTAAGCCTGAC 6453
Db 1804 --- GluThrAlaThrAlaLeuProGlnGlyGluGlyAlaAlaProSerHisSerArgAla 1822
QY 6454 AATGATATAAAGAGCGCTTTTATTATATGTCAGCAATAGAACTCAATTTTGTGCTTA 6513
Db 1823 AspAspLeuArgAsnAlaIleGlnSerAlaIleIleGluThrPhePheLeuTrpAsp 1842
QY 6514 AATATATATATGAA --- AATCTGCAGCAGAAAAATGAA --- 6549
Db 1843 ArgTyrLysGluGlnLysLysProGlnGlyAspGlySerGlnAlaLeuSerGlnLeu 1862
QY 6550 ----- TTGCAAAATGGAACA 6564
Db 1863 ThrSerThrTyrSerAspAspGluIleAspProProAspLysLeuLeuGlnAsnGlyLys 1882
QY 6565 ATTCAGATGAATTTAAAGAAATATATATATATATATATATATATATATATATATATAT 6624
Db 1883 IleProProAspPheLeuArgLeuMetPheTyrThrLeuGlyAspIleLeu 1902
QY 6625 TTTCGAACATGATTTCTTAATGATAAAAAATATATATATATATATATATATATATAT 6684
Db 1903 ValHisGlyLysAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1922
QY 6685 ATTCGATGAATATAATAGAAAAACAGATAAAAAAGATGAAGATTAACGTA 6744
Db 1923 ValLeuGluAlaSerGlyAsnLysGluAspMetGlnLysIleGlnLysIleGln 1942

QY 6745 ATA ----- 6747
Db 1943 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1962
QY 6748 ----- TTTCGGAGAAAAATAAAATTTATTTGGGAAGATGATATATGATTAATTTAT 6804
Db 1963 LysTrpTrpAsnGlnHisAlaGlnSerIleTrpLysGlyMetIleCysAlaLeuThrTyr 1982
QY 6805 ----- CATCTCACGAGCAAAACGAAAAAGAAAAATTA 6840
Db 1983 ThrGluLysAsnProAspThrSerAlaArgGlyAspGluAsnLysIleGluLysAspAsp 2002
QY 6841 GATTAATTAC ----- 6849
Db 2003 GluValTyrGluLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2022
QY 6850 ----- CAGTACAT --- GACATACCAAACTG --- 6873
Db 2023 ThrGlyThrTyrLysThrGlnTyrAspTyrGluLysValLysIleAspThrSerGly 2042
QY 6874 ----- ACCCTTCCTCGAAGAGCTTTGAAAAAGCGCC 6906
Db 2043 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheValLeuArgPro 2062
QY 6907 CAATTTTGAATGTTCAACAGATGGCAGAGAAATTTGTATATAGAGAAAGACAG 6966
Db 2063 ProTyrPheArgTyrLeuGluGlnIleTrpGlyGlnAsnPheCysLysAspArgLysHisLys 2082
QY 6967 TTGTTAAATTTGAGCGCGGCTGTAAAGAAATATGAG ----- 7002
Db 2083 LeuAlaGlnIleLysHisGlnCysLysValGluGluAsnGlyGlyLysSerArgArgGly 2102
QY 7003 ----- TGTAAAT ----- GGTAAAT 7017
Db 2103 GlyIleThrArgGlnTyrSerGlyAspGlyGluAlaCysAsnGluMetLeuProLysAsn 2122
QY 7018 GACGTAGACACACAGAA ----- TGTGACAGGCGCTGTATACATATCA 7062
Db 2123 AspGlyThrValProAspLeuGluLysProSerCysAlaLysProCysSerSerTyrArg 2142
QY 7063 AATTTATTAAGAAAGTGAAGAACTGAATATGAAGAAACAAAGAAAGTTCAAAAGAT 7122
Db 2143 LysTrpIleGlnSerLysGlyGluPheGlnGlyGlnLysLysLysLeuGlnGln 2162
QY 7123 AAAGAT ----- GGCAAAAGATTAAGCATTAATCTTCTCTGAAAGACACATA 7170
Db 2163 LysAspLysCysValAsnGlySerAsnLysHisAsp ----- 2174
QY 7171 GAGAAAGCAAACTGCTCATGAATATTAAACATGAATAAAGAAATATATGTCAT 7230
Db 2175 ----- Asn 2175
QY 7231 AAGGATGTTCTTGATGCAAAAACTTCTTCACAACTACCAAAAAACAACACACATCA 7290
Db 2176 GlyPheCysGlnThrLeu ----- Thr 2182
QY 7291 CATCATCCGATCGTAAT 7350
Db 2183 ThrSerSerLysAlaLysAspPheLeuLysThrLeuGly ----- 2195
QY 7351 AAGTGTAGTCCCTGAACTTTCAAAAAAGGATATATATATATATATATATATATATAT 7392
Db 2196 ----- ProCysLysProAsnAsnValGluGlyLysThrIlePheAspAspAspLysThr 2213
QY 7393 ----- CATCAAAAAAATAATTAAGAACT --- AAAATACCTATGATTTGTATAGAG 7440
Db 2214 PheLysHisThrLysAsp --- CysAspProCysLeuLysPheSerValAsnCys --- 2230
QY 7441 AAGCAGCATATATTATCTTAAGAAGCAAGAAATAATATGATATTACCTTGAAGAA 7500
Db 2230 ----- 2230

QY 7501 AAATTATACCTTTAGCTACAAAGAGAAAGTAAAAATAGTTGACTAATTAAT 7560
Db 2231 -----
QY 7561 AATCTTCGATCTAGAAACCTTAATGACCGATTAATATATAGAGAAACAAACCTT 7620
Db 2239 -----
QY 7621 TGTGAAATATGAGAGAAATCGTTTAAAGTAGATTATGAAATGTAATGTTCAAAAAT 7680
Db 2243 CysArgAsn-----
QY 7681 TCAAGTCTATACGAGAGAAAGAAAGTATGTGTACCTCCAAAGAGAACATATGTGC 7740
Db 2248 Ser-----
QY 7741 TTAAGAAATTTAGATGAAATTAATTAAGAAAGACTTAAGATGTAATTAATGCTTAATA 7800
Db 2249 -----
QY 7801 ATGCTTCGTGAACTGACGAGAAATGAGAAATGACATATATAAAACCTTCAACTCA 7857
Db 2266 ArgValSerAlaSerLysSer-----
QY 7858 -----GAGAACGGGTGCGCATGATCCAAATATGTGATCTATGAAATATAGTTGCT 7911
Db 2279 GlyLeuGluAsnAlaCys-----
QY 7912 GATCTGGTGACATAGTTAGAGAACAGATATGTTACGAATTTGGTGTACTTACCTGCC 7971
Db 2285 -----ArgGlyAlaGly-----
QY 7972 GTAGAAATTAATTAATTAAGCTTTTGAATACATATATGAAATGAGAAATATAAAT 8031
Db 2289 -----IlePheGluGlyIle-----
QY 8032 AAAGTAGAATAATAATACACGATGACAAAGCTTTCGTTGCTGGATGCTAAT 8091
Db 2293 -----
QY 8092 ACAAAAGATATTTGGAAGCA-----ATGACGTGCAGAAAGCCAGAA 8133
Db 2294 ArgLysAspGluTrpValGluTrpPhe-----PheGluAspTyrAsnLysIle 2344
QY 8134 GATGCAAAACTTTTGAAGAAAGAAAGATGATGATTTGAACGATACATTAATACAA 8193
Db 2313 AsnValAsnGlyLysIleLysGlyLys-----
QY 8194 GATAGGTGACATAGAGACATCCACCTGTTGATGATTAATATACCTCAACGGTTTGA 8253
Db 2322 -----HisIleIleGlnIleArgAla 2328
QY 8254 TGGATGACTGAATGCTCGAATATTAATTTGTAAGCAGTATGAGAAATTTG 8313
Db 2329 LeuValLysArgTrpValGluTrpPhe-----PheGluAspTyrAsnLysIle 2344
QY 8314 AAAAAATATGATGACATCTGTAAGAAACATCTGACAGATGCAAGATGATTAATGAGAAAT 8373
Db 2345 LysHis-----LysIleSerHisArgGlyIleLysAsn-----Gly 2355
QY 8374 AAGTGTGACAGTGTAAACGAGATGTCAAGATATATAATAATTTGTTCTTAATGAGAA 8433
Db 2356 GluIleSerProCys-----IleLysAsnCysValGluLysTrp--- 2368
QY 8434 TCTCTATTCGATATACATCAATAATATACAAAGAAATTTGATGAACAACCAATATATACA 8493
Db 2369 -----ValAspGlnLysArgGluTrpLysGluIleThrGluArg--- 2382
QY 8494 AAAATCTCTACTATGATCATGTTCAAAATTTTGTACAAAAGTTGAAACCTTTTAAAGT 8553
Db 2383 -----PheLysAsp 2385
QY 8554 GAATGTTCTGTGAGAGCTTTTCTGAATATCTCATGAACAGTAAGTGTAAATAT 8613

Db 2386 Gln-----Tyr 2387
QY 8614 AAATTATGAATAANGAGTCTTCCATATATACAGATATGCTTTGCAAGAACACCA 8673
Db 2388 Lys-----AsnAspAsnSerAspAspAsnValArgSerPheLeuGlnThrLeuIlePro 2406
QY 8674 AAA-----AGTTATTAAGAGCT 8691
Db 2407 GlnIleThrAspAlaAsnAlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSer 2426
QY 8692 TCCAGTTTACACTCTCTTCAGAAATCCATGATTAATTTGCTACGCAAAACAA 8751
Db 2427 CysGlyCysSerAlaSerAla-----AsnGluGlnAsnLys 2438
QY 8752 GATGATGTAGCAATTTCAAACTTTTACCTTCCTGCTGGAAGAAATGATTAATTAAT 8811
Db 2439 AsnGly-----
QY 8812 CTTGATTAATGAGCATACCTTGTCTTAATAGTTCAGATGATACAAAGTGTATTTG 8871
Db 2440 -----
QY 8872 ATTCCTCCAGAGAGACATTTATGTACAGACCTATCACTGCATATTAATTAAGAAA 8931
Db 2440 -----
QY 8932 GGTGATTAAGAAATTTTAAAAAACTTCTACTTCTGCTTCAGTTCAGAGCAATTTG 8991
Db 2440 -----
QY 8992 TTAGCTCAAAATTAATAATCGAAGAGAGTGTGCTTTGAGCAATGAATATAGTTAT 9051
Db 2440 -----
QY 9052 GCAGATTATTCGATATATTAATTAAGAACTGATATGATGACACTTATATGTAAGAAA 9111
Db 2441 --GluTyrLysAspAlaIle-----AspCysHleuLysLysLysLysAspLys 2456
QY 9112 ATTAATAAAATATTTGAACATCAATGAACCAACCAAAATCTGTAATAACATGCTGGGAA 9171
Db 2457 Ile-----
QY 9172 AATATAGCGTCAGATATGCGACGCTATGTATGTGATATAAATTTGCTACTCAAAA 9231
Db 2457 -----
QY 9232 GTAACTATGATGAGAGATGCTGTAATTAACAAAGATGAGAAACTAATCAGTTTCTT 9291
Db 2458 -----GlyGluCys----- 2460
QY 9292 CGTTGTTAATTTGATGAGGCAAGCAAGCATGTAAGAAAAGAAACATGTAAGTATTC 9351
Db 2461 -----GluLysLysHisLysGlnThrSer 2468
QY 9352 TTAATAAATAATGCT-----CTCGTTCAAAAGCAAGTAAATTTTGAAGG 9396
Db 2469 --AspHrcGluCysSerAspHrcProGlnThrLeuGlnLysP--- 2483
QY 9397 TCGAATTAATTAAGACAACTGATGATGATTAATTAATTAATTAATTAATTAATTAAT 9456
Db 2484 --GluThrLeu-----AspAspAspIleGluThrGluGluLysLys 2497
QY 9457 ATATGATTAATAAATAATGAAATATCTTAATTAATTAATTAATTAATTAATTAATTAAT 9516
Db 2498 AsnMetHrcProLysIleCysGluAsnVal---LeuLysThrAlaGlnGlnLysAspGln 2516
QY 9517 TCTCAGGTAATATAGACATAAACCATCTGAAGAAAGTGTACAGCATATATAAATCA 9576
Db 2517 --GlyGlyCysValProAlaGluAsnSerGluGluProAla-----Ala 2530
QY 9577 AAAGATTCTCAATGCGCTTTGAGATTAAATGATATTAATTAATTAATTAATTAATTAAT 9636

[illegible]

QY 1972 -----CAATCGTATTAGTAAATTAATTCCTTTTGAAGGTATTTT----- 2016
DB 736 Proheasph1sAspYrValLeuGlnTyrAsnLeuGlnGluLupheLeuLysGlyAsp 755
QY 2017 -----TTTAAAGTATGAGATTAACCTTCAACAAAGTGAAGCAAAATGCG 2058
DB 756 SerGluAspAlaSerGluGluLysSerGluAsnSerLeuAspAlaGluGluAla----- 773
QY 2059 AAGAACTTATGCAAAATATATAAAGAAAAAATGAGTTCATTCATTTGGAATAAT 2118
DB 774 -----GluGluLeuLysHisLeuArgLileLeuSerGluAspAsn 789
QY 2119 AGGAGC-----TATTTAGAGATGCAATAGACTCTGTTA 2154
DB 790 GlnGluAlaSerValGlyGlyValThrGluGlnLysAsnIleMetAspLysLeu 809
QY 2155 GATCACTTAAGAACTGCGCAGATATGT----- 2184
DB 810 AsnTyrGluLysAspGluAlaAspLeuGluLileHisGluAspGluGluGlu 829
QY 2185 ---AAGACATATATACAAAGCAAGCATGCAACATCCATATGCAACAAACCCG 2241
DB 830 LysGluLysGluLysAspGluLysGluLysIleGluGluLysGluAsnPheArgTyrAsnPro 849
QY 2242 TGTGTTAACTCTGCGAGCAGCAAA---CCCACTAAAAATATAAAGAAATAGCAAA 2298
DB 850 CysSerGlyLysSerGlyAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGln 869
QY 2299 TACTTTAAAGAGTGCATACGAGAAAGCAAGCAATCGTGTCTCTCAATATGAAAGA 2358
DB 870 MetHisHisLysAlaLysThrGlnLeuAlaSerArgLileLysSerAlaLeuGly 889
QY 2359 AAGGACACGAGATATATATAAAGCTGGGAGTGAAGAAAGCACTCAAGCAATTTA 2418
DB 890 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGluGlnIle 909
QY 2419 TGTGATATATGATTAACATTTCTATCTGATCTGTTTCAATGACACATGAT 2478
DB 910 CysLysLysLeuGlnLysAsnTyrSerAsnAspSerArgLysSerGlyCysProCysThr 939
QY 2479 GCGAAAGCAGCAGGTGATATACAAAGAAATTTGCTAGAACTGAAAGGAAAGT 2538
DB 930 GlyLys---AspLysPheLysGlyValArgMetArgLileGlyThrGluTyrPheAsn 948
QY 2539 GATCCGGAACACATGCGTAAAGATCAAGAGATGTTATGCTCTAGAACGACAT 2598
DB 949 IleGluGlyLysLysGlnThrSerTyrLysAsnValPheLeuProProArgArgGluHis 968
QY 2599 ATATGATCATCAATTTGGAACATTTACAAAGGAGATGACCCACCTTAATGATATAT 2658
DB 969 MetCysThrSerAsnLeuGluAsnLeuAspVal-----GlySerVal 982
QY 2659 GTT---GATGATTTAGTATTAATTCCTTTTGGGGAGTGTCTCTACAGCAAAATAT 2715
DB 983 ThrLysAsnAspLysAlaSerHisSerLeuGlnLysAspAlaGlnLeuAlaAlaLysThr 1002
QY 2716 GAAGCAAAACAGATATATACATGATATTAAGAAAGATTAACCTAAGGGCCCAAGAA 2775
DB 1003 AspAlaAlaGluLileLysArgTyrLysAspLysAsnIle-----Gln 1018
QY 2776 GTAATGACCA-----AAACACAGACAACTATGCTGAGCTATACGTTACGT 2826
DB 1019 LeuThrAspProLileGlnLysAspGlnLysAlaMetCysArgValAlaArgTyrSer 1038
QY 2827 TTTGCAATATAGTATATATATTCGAGAAAGATCTCTGGGAAAGAAAGCGGACATG 2886
DB 1039 PheAlaAspLeuGlnLysPheLileLysArgLysArgPheTyrAspGluAspLysSer 1058
QY 2887 GTAACGCAAGACATTTGGAACCTTTTGGTATATATACATGACCTCAAGCGC 2946
DB 1059 ThrAspMetGluThrArgLeuLileThrValPheLysAsnLileLysGluLysHisAspGly 1078

QY 2947 ---AAGGAAATGATTAATTAATGATGAT-----GCCCCCAAAATTTTAAATG 2994
DB 1079 IleLysAspAsnProLysTyrThrGlyAspGluSerLysTyrProAlaTyrLysLeu 1098
QY 2995 AGGAAATATGTTGGGAGCTATATAGCCCAAGTATGGAACCGAATGAAATGATAT 3054
DB 1099 ArgAlaAspTyrPheGluAlaAsnArgHisGlnValTyrArgAlaMetLysCys----- 1116
QY 3055 AAATATTTGAGAAATTAATCGGACACCAATCAACAAAGTATGTTATTCGATATAGT 3114
DB 1117 -----AlaThrLysGlyLileLys----- 1123
QY 3115 GATCATACACCATGATGATATATATCCACAAATTAAGATGATGACCAATGCGCA 3174
DB 1124 ProGlyMetProValAspAspTyrIleProGlnArgLeuArgTyrPheMetThrGluTyrAla 1143
QY 3175 GAATGGTACGCAAGGTCGCAAGAAAGGAGTATGATGATGAAAGCAAGCTAAGAGC 3234
DB 1144 GluTyrPheCysLysAlaGlnSerGlnLysTyrAspLysLeuLysLysIleCysAlaAsp 1163
QY 3235 TGTAAAGATAGGATATATGCTCAAGGCTGACAAAGAGAGTGTACAGGTTGTACGAAG 3294
DB 1164 CysMetSerLysGluLysAspGlyLys---CysThr---GlnLysAspValaLysCysGlyLys 1181
QY 3295 TGCACAGACCTTGTAT 3354
DB 1182 CysLysAlaLysLysAspLysTyrLysGluGluLileGluLysThrPheAsnGlnTyrArg 1201
QY 3355 ATATATGAT 3414
DB 1202 LysIleSerAspLysTyrAsnLeuLeuTyrLeuGlnAlaLysThrThrSerThrAsnPro 1221
QY 3415 GGTATGGAACCTTCAGTACTGCGCAAAATCATATAGAC-----AGAAATGATAT 3465
DB 1222 Gly-----ArgThrValLeuGlyAspAspAspProAspTyrGlnGlnMetVal 1237
QY 3466 GAATTTTGTGCGAAT 3513
DB 1238 AspPheLeuThrProLileHisLysLysAlaSerIleAlaAlaArgValLeuValLysArgAla 1257
QY 3514 AGTGTACTGATGATGAAATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3573
DB 1258 AlaGlySerProThrGluLileAlaAlaAlaAlaProLileThrProLysThrAlaAla 1277
QY 3574 GCATATCTCCATGATACAGAAATTTGATGATGATGATGATGATGATGATGATGATGAT 3633
DB 1278 GlyTyrLileHisGlnGluLileGlyTyrCysLysLysGlnGlnPheCysGlu 1297
QY 3634 GAAAAA-----AGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3675
DB 1298 LysLysHisGlyAlaThrSerThrSerThrThrLysGlnLysGlnLysGluTyrThrPheLys 1317
QY 3676 GATAAACACAGGACCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3735
DB 1318 GlnProProGluTyrAlaThrAlaCysAspCysLysLysAsnArgSerGlnThrGlu--- 1336
QY 3736 GTACAGATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3795
DB 1337 -----GluProLysLysLysGluLysAsnValGlnSerAlaCysLysLileValGlu 1353
QY 3796 GATATACTTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 3855
DB 1354 LysLileGluGlnLysLysAsnGlyArgThrThrValGlyLysAsnProLysGluSer 1373
QY 3856 AGTATGATATATCCGATGCGCAATGC---GGAATATATATATATATATATATATATAT 3912
DB 1374 -----TyrProAspTyrPheAspCysLysAsnAsnLileAspLileSerHisAsp---Gly 1389
QY 3913 GTGTATATGCCCCCTAGACAAAGAAAGTATGCTATATATATATATATATATATATATAT 3972
DB 1390 AlaCysMetProProAspArgGlnLysLeuCysLeuTyrLileAlaHisGlnSerGln 1409
QY 3973 ATAAAAAATTTACATCAAGATTAATTTAAAGAGCTTTCATCAATCTGCAGCAGCA 4032

Db	1410	ThGlaaSnllelysthraspaSpasnuLeuLysaspAlaPheIleLysThAlaAlaIa	1429
QY	4033	GAACACTTCTTCATNGGATATATATATAAAGTAAAG---GATGTGAAGAAATGACATC	4089
Db	1430	GLUTHrPheLeuSerTribInTyRlyLysSerLysAsnAspSerGluAlaLysIleLeu	1449
QY	4090	GATTAAGAATTTAAAGAAAGCCAAATATCCCTCCGCAATTTTGGATGTCATGTGTACACA	4149
Db	1450	AspArg-----GlyLeuIleProSerGlnPheLeuAlaArgSerMetLeuTyThr	1465
QY	4150	TTTGGAGATATATAGCATTTTATTTATTTGGAACAGATATATATCAAAAGGTGATGTGAGGGA	4209
Db	1466	PheGlyAspTyAlaArgAspIleCysLeuAsnThrAspIleSerLysGlnAsnAspVal	1485
QY	4210	AGTAACATCAAAAGACCAATATGATTCCTTTTCAAAAATGGTGACCAAAAATCTCTAAT	4269
Db	1486	AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspGlySerLysSerProSer	1505
QY	4270	GGAATAACACCCCAAGCAATGGTGGACAGACAMCATGTCATGATATGGGAAGCATCTCA	4329
Db	1506	GlyLeuSerArgGlnGlnIuTPRlyThrAsnGlyProGluIleTyrPlySgIleMetLeu	1525
QY	4330	TGTGCACATGATTAATAATTTGGGGCAAAAAAGATGAT-----TTTACCAGAAACTAC	4380
Db	1526	CysAlaLeuThrLysTyRlyValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr	1545
QY	4381	GGTTCACAACAACGTCAAAATTTAGTACAAA---AGCACCCCTTTGGAGAAATTTGGCAAA	4437
Db	1546	SerTyrAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnIlePheAlaIa	1565
QY	4438	CGACCCCAAGTTTACATGGCATGCCATACCGAATGGTACGACATTTTGTATATACAGACAA	4497
Db	1566	LysProGlnPheLeuArgTyrPheMetIleGluTyrGlyGlnGluPheCysAlaGluArgGln	1585
QY	4498	AAATATTTTGAAGATGTGCAGAAAAATGTAACTCA---AATGACCAATTTGAAGTGTGAT	4554
Db	1586	LysLysGluAsnIleIleLysAspAlaCysAsnGlnIleAsnSerThrGlnIleCysAsn	1605
QY	4555	ACA-----GAATGTATATGAATAATGCGAGGACTACGTTAAATATATGAAA---AAA	4602
Db	1606	AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyRlyGlnIuTyRlyAlaGlnAsnLys	1625
QY	4603	AAAAAGAGTGCATGTCCACAAGATAATATTTCCAAAGAGATGACGGACAAAAAAGATTC	4662
Db	1626	LysLysGluPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln	1645
QY	4663	GATAGCAACAACATTTGGTGTAAATGGTTACAGACTATACGTGAACGAATGCAACAGATTAC	4722
Db	1646	AspProGlnIuTyRlySgIlyTyRlyGluTyRlyAsp-----	1656
QY	4723	TTGACACAGAAATTTACTGCTAGTGTGTGTGATATAGCTCGAAGTGCCTGTGTGTACAA	4782
Db	1657	-----GlyAlaGlnPro-----IleGln	1662
QY	4783	ACAATATATACAAATGTGTAGAAAAACAGCGCTTACTATGATGGCGACAACATTTGGGGTGC	4842
Db	1663	GlyAsnGluTyRlyLeuLeuGlnLys-----CysAsnAsnAsnLys--CysSerCys	1678
QY	4843	ACAAAATTTATTTGAANAATGACGACAAATATATACTAACAATTTGCACTAAAGATTAAGTCAAA	4902
Db	1678	-----	1678
QY	4903	GGATTAGTAAAGAGGCAACACAGGTGCTATTAAGTGGCAAAAACAAAGTCTTAATTAAC	4962
Db	1678	-----	1678
QY	4963	TACATATACTTGAAGAAGATTGACTGAAGAAGTGTCTTTTCTCTCTCGACTACGTATA	5022
Db	1678	-----	1678
QY	5023	TGTTTTCATGCATGATGAGTCAATTTATACAGATCCACAATTTAAAGTGAATAATGGCTTG	5082

Db	1679	-----	netaspelysnv\l1euser\valser\ProLysoLys-	-----	1691
QY	5083	CGAAAAAGATTGATGGAGTGGCGGCACGGAAGGATACAAATTGGTCAACTACTACAA			5144
Db	1692	-----	prophelylystyr-----	:	1696
QY	5143	GAIAAAAAAAAAAGAAAGAAAAATAAAAAGCTGGATCGGCACCAAAATTTCTTAATGCGTC			5202
Db	1697	-----	Alahlslystyr-Pro-----		1701
QY	5203	CCGCCCTGTAGTCTATGATAATATAGTTTTTATGATTTAAGAGATATATCTAGTATT			5262
Db	1701	-----			1701
QY	5263	GATTAATTGGAAATGAAAAACAAAAGACCGAGAAAATTTGAAGMAAAATTTTAACAA			5322
Db	1701	-----			1701
QY	5323	AATGACATCAGTTGGCAAAAGAGTGAATAGTACTACAGSAAATCCCGGTAGTACGCG			5382
Db	1701	-----			1701
QY	5383	CGAAATTTTCTGGAACGAAAAATAAGAAATGTGTGGAACGCAATGATATCGGGTAC			5442
Db	1701	-----			1701
QY	5443	AAAGGTGGTAGGATGATGAAATATAGGAATATGTGCMAAGAGTGAATGATCTAAA			5502
Db	1701	-----			1701
QY	5503	AAATGTGTTCTGTACTTACATGATGANTTACTATGGGAAAAATCCGATGAAGGT			5562
Db	1701	-----			1701
QY	5563	ACTCGTATCAGTTCTTCCATGGTTTCCGGAATGGGGTGAACATTTTGGCAACATAAA			5622
Db	1701	-----			1701
QY	5623	GAIAAGAAATTGGAAATTTGGTAGGGCGTGAATGATTACTTGTGGTGAATGA			5682
Db	1701	-----			1701
QY	5683	GATAAAAGAAAGATGTACAGATGGCGTGCACAAATATAAAAATTTATAGTAGTGG			5742
Db	1702	-----	glulyscys-----		1704
QY	5743	AAACCAAGTATGAAAAACAAATCAAAAAATATGTGAGAAATAAGACAAAAATATTCC			5802
Db	1704	-----			1704
QY	5803	GAGCATCTGTGCAAAAGATGACAGAGCGCTCGGAATATTAGACAAACAAATTA			5862
Db	1704	-----			1704
QY	5863	AAAAATTTGTGAATATMAAAAGTGAGATTGTGAATATTAAGTGAATAAAGATGTCCACA			5922
Db	1705	-----	aspcys-----		1706
QY	5923	CAGCATTAAGTGAATAGTATAGTCAAAATATATGCCGATCATTAAGACATGAACCA			5982
Db	1706	-----			1706
QY	5983	GAAGTTGAAGAAAGTGAATTTGCAATGCCACAGAGTCCACCACTGTACGAAGGAA			6042
Db	1707	-----	tyringlyls-----		1717
QY	6043	AAACGTCACACGGGTATCACTGATATCAAAAGCGACGCGATGAAAAAAGAACGAAA			6102
Db	1717	-----			1717
QY	6103	ACAGCGCGCTACAAAACAGCGAAAAAGTGAATAATCTAACACAGAAATGCGGCA			6162
Db	1718	-----	ProProProProvalGlnPProGln-----		1725

[illegible]

Db 790 GlnGluAlaSerValGlyGlyGlyValThrGlnGlnLysAsnIleMetAspLysLeu 809
QY 2155 GATCACTTAAAGAAATGCGACGATATGT----- 2184
Db 810 AsnTyrGlnLysAspGlnAlaAspLeuGlnIleHisGlnAspGlnGlnGln 829
QY 2185 ---AAAGCAATATATACAAAGCAAGCATGTGAACATCCATTAATGCAACAAACCCG 2241
Db 830 LysGlnLysGlnLysAspGlnAsnGlnCysIleGlnGlnGlnLysAsnPheArgTyrAsnPro 849
QY 2242 TGTGTAAACCTCGTGGAGCAGCCAA---CCCACTAAATAATTAAGAAATAGCAACA 2298
Db 850 CysSerGlnLysGlnSerGlnLysAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGln 869
QY 2299 TACTTTAAAGAGTGCATACGAGGAACAGCAAAATCGTGCTCTGATTAATTAAGGA 2358
Db 870 MetHisHisLysAlaLysThrGlnLeuAlaSerArgAlaGlyArgSerAlaLeuArgGly 889
QY 2359 AAGCAACAGAGGTATATATTAACGTGGCGGTAGAGAAAGACATTCAAGCAATTTA 2418
Db 890 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlnIle 909
QY 2419 TGTAGATTAATGATTAACATTCATTAATCGTATCTGTTTCAATGAGACATGTAT 2478
Db 910 CysLysIleAsnGlnLysAsnTyrSerAsnAspSerArgGlyAsnSerGlnLysProCysThr 929
QY 2479 GGCAGAGCAGCAGTGTATATACAAACAGATTTGCTGTAGGAATGAGTGGAGT 2538
Db 930 GlyLys---AspGlnLysAspHisGlnGlyValArgMetArgIleGlyThrGlnTyrSerAsn 948
QY 2539 GATCCGGAACACATCGGTAAAGATCAGAAAGTATATATATATGCTCTGAGACACAT 2598
Db 949 IleGlnGlnLysLysGlnThrSerTyrLysAsnValPheLeuProArgTyrGlnLys 968
QY 2599 ATATGTACATTCATTTGGAACATTTGAACCGATGATCACCACCTTAATGATATAT 2658
Db 969 MetCysThrSerAsnLeuGlnLysAsnVal-----GlySerVal 982
QY 2659 GTT---GATGATTTAGTAAATTCCTTTTGGGGGATGCTCTCTATACCAAAATAT 2715
Db 983 ThrLysAsnAspLysAlaSerHisSerLeuGlnLysAspValGlnLeuAlaLysThr 1002
QY 2716 GAAGCAACAGATTAATGCAATGTATAAGAAAGAAATTAACCTTAAGGCCCCAAGAA 2775
Db 1003 AspAlaAlaGlnIleLysArgTyrLysAspGlnAsnAsnIle-----Gln 1018
QY 2776 GTAACGTGACCCA-----AAACACGACAACTATCTGTGAGCTATACGTTACAT 2826
Db 1019 LeuThrAspProIleGlnGlnLysAspGlnLysAlaMetCysArgAlaValArgTyrSer 1038
2827 TTTCGAGATATAGTATATATTCGAGAGAGATCTCTGGGAAAGAAACGCTACATG 2886
Db 1039 PheAlaAspLeuGlnLysPheLysArgGlnAspMetThrAspGlnAspLysSerSer 1058
QY 2887 GTAAGCTGCAAGACATTTGGAAAATCTTTTGGTAAATATACAAATGCTACAAAGC 2946
Db 1059 ThrAspMetGlnThrArgLeuLysPheLysAsnIleLysGlnLysHisAspGly 1078
QY 2947 ---AAAGGAATGATAATTAATATATATAT-----GCCCCAAATATTAATATG 2994
Db 1079 IleLysAspAsnProLysTyrThrGlnLysAspGlnLysProAlaTyrLysLysLeu 1098
QY 2995 AGGGAATAATTTGGTGGAGTAAATAGACCAAGATGAGGAAGCAATTAATGATATA 3054
Db 1099 ArgAlaAspTyrPheGlnAlaAsnArgHisGlnValThrPheAlaMetLysCys----- 1116
QY 3055 AAATATTTTGAAGATAAATCGGACACCAACACAAAGTATATTTGGCGATATAGT 3114
Db 1117 -----AlaThrLysGlnLysIleLysCys----- 1123
QY 3115 GATCATACCATTTGATGATTTATATCCCAAAATTAATAGATGATGACCAATGGGCA 3174
Db 1124 ProGlnMetProValAspAspTyrIleProGlnArgLeuArgTyrMetThrGlnTyrAla 1143

QY 3175 GAATGCTACTCGAAGGTGCAAGAAAAGCATATGATTAAGTTGCAAGAGATGTAAAGAG 3234
Db 1144 GlnTyrPheCysLysAlaGlnSerGlnGlnLysTyrAspLysLysLysIleCysAlaAsp 1163
QY 3235 TGTAAAGATTAAGGATTAATGCTCAAGGCTGTACGAAGAGTGTACAGTTCACAG 3294
Db 1164 CysMetSerLysGlnLysAspLysLys---CysThr---GlnGlyAspValAspCysGlyLys 1181
QY 3295 TGCACAGAACTTGTAAATGATTAATATATATATATATATATATATATATATATAT 3354
Db 1182 CysLysAlaAlaCysAspLysTyrLysGlnGlnIleGlnLysTyrPheAsnGlnThrArg 1201
QY 3355 ATAAATGATGATTAATTAACAAAGATTTACATGACACAGCAAAATGCTGTATATAGT 3414
Db 1202 LysIleSerAspLysTyrAsnLeuLysLeuGlnAlaLysThrThrSerThrAsnPro 1221
QY 3415 GGTATTTGAAGCTTCCAGTACTGCCAAATAATCATATATAGAC-----AGGAATGTTAT 3465
Db 1222 Gly-----ArgThrValLeuGlnLysAspAspProAspTyrGlnGlnMetVal 1237
QY 3466 GAATTTTGTGCGAATTTATACCAACAAATGTGGCAAA-----AGTAATATA 3513
Db 1238 AspPheLeuThrProIleHisLysLysSerIleAlaAlaArgValLeuValLysArgAla 1257
QY 3514 AGTGTACTAGTGAAGAGTGTGCTCATTTGCTACTAACACACAGTATGAAATGTTGCA 3573
Db 1258 AlaGlnSerProThrGlnIleAlaAlaAlaAlaProIleThrProTyrSerThrAlaAla 1277
QY 3574 GCATATCTCCATGATATACAGAAATTTGATGATGTGTCAGTCCAAATGAGTTTGTAT 3633
Db 1278 GlyTyrIleHisGlnGlnIleGlnIleGlyTyrGlnLysGlnGlnGlnGlnPheCysGln 1297
QY 3634 GAAAA-----AGTATGTTAAGATTAACGAAAAATATGCTTTAGA 3675
Db 1298 LysLysHisGlnLysAlaThrSerThrSerThrLysGlnLysAsnLysGlnTyrThrPheLys 1317
QY 3676 GATTAACCAACAGCATGATGATGCTGCTGCTGTTGTTAAAGTGTGGAACCGACAG 3735
Db 1318 GlnProProGlnTyrLysAlaThrAlaCysAspCysLysAsnArgSerGlnThrGln 1336
QY 3736 GTACAGATTAATAAGCAAAAAAGCGGAAAGAAAGATGCGAATGTAAACAGTGAAT 3795
Db 1337 -----GlnProLysLysLysGlnLysGlnLysValGlnSerAlaCysLysIleValGln 1353
QY 3796 GATATACTTAAGAAAGCATGAGTGAAGAAACAAAGTGAACATTTGCAATTAATGA 3855
Db 1354 LysIleLeuGlnGlnLysAsnGlnLysArgThrThrValGlnLysCysAsnProLysGlnSer 1373
QY 3856 AGTATGATATCCGATTTGGCAATGC---GGAATATTAATTTAGTGAAGACCTCGT 3912
Db 1374 -----TyrProAspTyrAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1389
QY 3913 GTGCTATGCCCTGAAACAAAGATTTGCTGATATTTCTTGCAATTAATGA 3972
Db 1390 AlaCysMetProProArgAlaGlnLysLysCysLeuTyrTyrIleAlaHisGlnSerGln 1409
QY 3973 ATAAAAATTAACATACAAAGTAAATTTAAAGAGCTTCAATTCGAGAGCA 4032
Db 1410 ThrGlnAsnLysThrAspAsnLeuLysAspAlaPheIleLysThrAlaAlaAla 1429
QY 4033 GAAACATTTCTTCATGATTAATTAATTAAGTAAAG---GATGTTGAAGCAATGAACTC 4089
Db 1430 GlnThrPheLeuSerTyrGlnTyrLysSerLysAsnAspSerGlnAlaLysIleLeu 1449
QY 4090 GATTAAGAAATTAATAAGCAAAATTCCTCCCGCATTTTGGATGATCCATGTTACACA 4149
Db 1450 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetMetTyrThr 1465
QY 4150 TTTCGAGATTAATGATTTTATATTTGACAGAGATATCAAAAGTGTATGTAAGGA 4209
Db 1466 PheGlnAspLysArgAspIleCysLeuAsnThrAspIleSerLysGlnAsnAspVal 1485

Qy	7063	AAATTTATTAACAACTGGAAAACTGCAATATGCAAGAACAAAGAAAAAGTTCAAAAAGAT	7122
Db	2145	LYSTPLIEGLISerLysGlyLysGluPheGluIleGlnGlnLysAlaIleGluGln	2164
Qy	7123	AAAGAT-----	7128
Db	2165	LYSASPILysCysLysAlaSnGlySerAsnLysHisAspSnGlyPheCysGluThrLeuThr	2184
Qy	7129	---GGCAAAAGATATAGGATTAAT-----CCTTCACAGAAAGACATA	7170
Db	2185	ThrSerSerLysAlaLysAspPheLeuLysThrLeuGlyProCysLysProAsnVal	2204
Qy	7171	GAGAAAGCAACGTCGTCATGCAATATTTAAACATGAATTTAAAGATATATGCGCAAT	7230
Db	2205	GluGlyLysThrIlePheAspAspLysThrPheLys-----HisThr	2219
Qy	7231	AAGGATGTTCTTGATGCAAAAACCTTCACAACTCCAAAAACAACACAAATCA	7290
Db	2220	LYSASPcysAspProCysLeuLysPheSerValAsnCysLysLys-----	2234
Qy	7291	CAATCATCGAATGCAATGATATGCCGAATCGCTGATTAATGTTCTCGAAGATTTAAC	7350
Db	2235	-----Asp	2235
Qy	7351	AAGTGTGAGTCTCCTGAACCTTTCAAAAAGGATCTATGATTCATACAAAAAATTACT	7410
Db	2236	GLUCysAsp-----AsnSerLysGlyThrAspCysArgAsnLys-----	2248
Qy	7411	GAACCTAAATACCTATGATATGTGTGATGAGAAAGCAGCATATATTTATCTTAAAGAACGA	7470
Db	2249	-----AsnSerIleAsp-----AlaThrAspIle	2256
Qy	7471	GAATTAATATGATATTAACCTTGAAAGAAAAATTTATACCTATGAGCTACAAAGAA	7530
Db	2257	GluAsnGlyValAspSerThrVal-----LeuGluMetArgValSer	2270
Qy	7531	AAGGAAGTAAATATATGTGGACTATATATATCCTTGCGATCCTAAGAAACCTTATGCA	7590
Db	2271	AlaAspSerLysSerGlyPheAsnGlyAsp-----	2280
Qy	7591	CTGTAAATATATAGCAAGAACAAACCTTGTGAAAT-----	7629
Db	2281	-----GlyLeuGluAsnAlaCysArgGlyAlaGlyIlePheGluGlyIle	2295
Qy	7630	AGAGCAAAAATCGTTTAAAGTAGTATTTGATGATGAAATGTTACMAAATTCMAAGTTC	7689
Db	2296	ArgLysAsp-----GluThrLysCys-----ArgAsnValCysGly	2307
Qy	7690	TATCAGAGAAAAAAGAGATGTGACCTCCAAAGAAACATATGTGCTTAAGGAAT	7749
Db	2308	TyrVal-----ValCysLysPro-----	2313
Qy	7750	TTATAGTAAATTTAAATTTGAAAGACTTAAGATAGTAATATCTCCAAAAATGTTTCTG	7809
Db	2313	-----	2313
Qy	7810	CGAAGTCAAGAAATAGGAATAGACATATTAATAAACTCAGCTCAGAAACGGGTGC	7869
Db	2314	-----GluAsnValAsnGlyLeu-----	2319
Qy	7870	GCAATGATCAATATGTGATCTATGAATATATAGTTTCGCTGATCTGGGTGACATAGTT	7929
Db	2320	-----Ala	2320
Qy	7930	AGAGCAACAGATATGTTGCAATGTGGTGTACTTACCTCCCGTACGAATAAATTTATAT	7989
Db	2321	LysGlyLysHisIleIleGlnIleArgAlaLeuValLys-----ArgThrVal	2336
Qy	7990	AAGGTTTGTGAATACATATATGAAATAGGAATAGCAATATAAAGGTAGCAATATATAC	8049
Db	2337	GluThrPhePheGluAspArgLysAsnLysIleLysHisLysIleSerHisArgIleLysAsn	2356


```

QY 8050 AACGAT-----GTACAAAGCTTCTGCTGCTGGGATGCTAATAGAAA 8097
Db 2357 GlyIuILserProCysIleValGluLysTyrValaspGluLysArgLys 2376
QY 8098 GATATTGGAAGCAATGACGTCAAGACACGAGATGCAAAACTTTTGAAGAA 8157
Db 2377 Glu---TrrpLysGluIleThrGluArgpHeLysasp----- 2387
QY 8158 AGAATGATGATTTGACGCATTAACATTAAATACAGATAGTGTGACATTAAGACGAT 8217
Db 2388 -----GlnTyrLysAsnAspAsnSerAspAsp 2396
QY 8218 CCACCTGTGATGATATATATACCTCAACGTTTCGATGATGACGTAAAGTGTGATAT 8277
Db 2397 AspAsnValArgSerPheLeuGluThrLeuIleProGlnIleThrAspAlaAsn----- 2414
QY 8278 TATTGTAAGCACTGATGAGATGGAATTTAAAAATCATGTCATGCTGATAA 8337
Db 2415 ---AlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSerCysGly---CysSer 2432
QY 8338 ACATCTGACAGATGCAAGATGATTTATGATGAATAATAGTGTGACAGTGTAAACGAGA 8397
Db 2433 AlaSer-----AlaAsnGluGlnAsnLysAsn----- 2441
QY 8398 TGTCAAGATATATAAATTTTCTTAAATGGAATCTCTATTCGATATACATCAAT 8457
Db 2442 ---GlyLyuIrrLysAspAlaIle-----AspCysMetLeuLys 2453
QY 8458 AATATACAAAGATTTGTATGACAAACAAATATACAAAAATCTCTACTATGATCATGTT 8517
Db 2454 LysLeuLysAsp-----LysIleGlyLysGluLysGly 2465
QY 8518 CAAATTTGTACAAAGTTGAAACTTTTAAAGTGAATGTTCTGTGTGACAGCTTTTCT 8577
Db 2466 HIsHIs-----GlnThrSerAspThrGluCysSerAspThrProGlnPro 2480
QY 8578 GAATATCTTCATGAACAAAGTAGTGTGATTAATTAATTAATGAATATGATGTTCT 8637
Db 2481 GlnThrLeuGluLysp-----GluThrLeuAspAspIleGluThrGluGluAlaLys 2498
QY 8638 TCCAAATATACGAATATGCTTTCGAAGAAACCAAAAGTTATTAAGAA----- 8688
Db 2499 LysAsnMetMetProLysIleCysGluAsnValLeuLysThrIleGlnGlnLyspGlu 2518
QY 8689 ---GCTTGCAGTGTGACACTACCTCTCAAGATTCATGGATTAATGTCCTACCGATCAA 8745
Db 2519 GlyIuLys-----ValProIleGluAsnSerGluLysProIleAlaIleThrAspSer 2535
Db 8746 AACAAAGAT 8754
Db 2536 GlyLysGlu 2538

```

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fullier, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1PMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-12

Alignment Scores:
Pred. No.: 1,086-217 Length: 2710
Score: 2972.50 Matches: 865
Percent Similarity: 40.06% Conservative: 442
Best Local Similarity: 26.51% Mismatches: 873
Query Match: 15.28% Indels: 1083
DB: 4 Gaps: 120

US-10-087-013-1 (1-10628) x US-09-210-288-12 (1-2710)
QY 25 ATTATATAAATGGGAGATGACATGATCATTAAGAGAGATGCTAAAGCCCTATTATA 84
Db 6 LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspIleGluLysp----- 22
QY 85 AAGAAGTGCACAAAGTGCAGAAATGTTTGAACGTTATGCCAAATATTAAGCAT 144
Db 23 -----GlnSerIleLysHisMetPheAspArgIleGlyLysAspValIrrAsp 38
QY 145 CCATCAAAA---TATGCAAAAGACATGTGATTCGTGAAAGGGATTTGACGAAGACA 201
Db 39 LysValLysGluGluAlaLysGluArgGlyLysGlyLeuGlnLysArgLeuSerGluAla 58
QY 202 GAATTTGCTGTGTGCTTCTACGCCAGTAATTAAGCATTAATTAATTAATTCATATGCA 261
Db 59 LysPheGluLysAsnGlnLysSerProGlnThrProGluLysp-----Pro 73
QY 262 TGTAAATTTAGATCATTAAGACATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 321
Db 74 CysAspLeuAspHisLysIrrHisThrAsnVal-----ThrThrAsnValIleAsn 90
QY 322 CCTTGCATGCTAGAGACAAACCGATTTGATGAAGATCAATGTAATGT---GGA 378
Db 91 ProCysAlaAspArgSerAspValArgPheSerAspGluTyrGlyLysIrrHis 110
QY 379 AATTAATATGATTAATTAATAAAGAAATTAATGATGATTAATGATGATGATGATGATGAT 438
Db 111 AsnArgIleLysAspSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 130
QY 439 CAGCATATGTCATTAATAAAGCTGGAAGCTGTAATGAATGAATGAATGAATGAATGAAT 498
Db 131 LeuHisValCysAspGlnAsnLeuGlnIleGluProIleLysIleThrAsnThrHis 150
QY 499 GATTTATGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558
Db 151 AsnLeuLeuValAspValCysMetAlaIleLysPheGluGlnIleThrGlnAsp 170

```

```

OY 559 CATCA---CATAGAGACT-----TCAGACGGCTTGACTGCTTCA 600
Db 171 TyrProLysTyrGlnAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 190
OY 601 CGAATTTTGGCATATAGGTGATTTGTAAGAGATAGATATGTTTAAACCAATGTC 660
Db 191 ArgSerPheAlaAspIleGlyAspIleValArgGlyArgAspLeuTyrLeuGlyAsnPro 210
OY 661 CATGAC-----AAAGTAGAAGGGGTCCGACAGCGTTTCAACAATA 705
Db 211 GlnGluIleLysGlnArgGlnGlnLeuGlnAsnLeuYsnThrIlePheGlyValLe 230
OY 706 CATGAT-----GGATGGAAGATGAACTAAAAATGATTCATCTGATGATCT 756
Db 231 TyrGluLysLeuAsnGlyAlaGluAlaArgTyrGlyAsnAsp-----ProGlu----- 246
OY 757 GGAATTTTATTAATTAAGAGACATGCGTAATGTAATGATAATTAATGAGAA 816
Db 247 -----PhePheLysLeuArgGlnAspTrpTrpThrAlaAsnArgGluThrValIleTrpLys 264
OY 817 GCTATAACATGTCATGTCATTAATCTGATATTTTATGCAATGCAAGATTAATACA 876
Db 265 AlaIleThrCysAsnAla---TrpGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 282
OY 877 CCATATTTCGAATCCTTAATCGGCCATTAACAGAAAGGTTCCTACCAATTATAGT 936
Db 283 GlyGluArgThrLysGlyTyrCysArgCysAsnAspAspIleValProThrTyrPheAsp 302
OY 937 TATGTCCTCAATATTTACGTTGCTGTTGACAGAGAGAGAAAGTTTCCCAAAAAGA 996
Db 303 TyrValProGlnTyrLeuAlaGlyTrpPheGlnGluTrpIleAlaAspPheCysArgLysLys 322
OY 997 AATATTAAATGAAAAAGTCAAGACTCCTGCTG-----AATGCAAAAGAA 1044
Db 323 AsnLysLysIleLysAspValLysArgAsnArgGlyLysAspLysGluLysAsp 342
OY 1045 CGCTTATATATGTGATCATATGATGACATGTGACACAACTTTGGAAGAAAGTATT 1104
Db 343 Arg---TyrCysSerArgAsnGlyTyrAspCysGluLysThrLysArgAlaIleGlyLys 361
OY 1105 TTGCATTTGGATTAATAGTGTACTGACTGTGCACTTAATGCAAGATTGTAAGTTGG 1164
Db 362 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTrp 381
OY 1165 TTAGGAATCAACAAGACATTTAAACAAACAAAGAAATATGAAAAAGAAATACAA 1224
Db 382 IleAsnAsnGlnLysGlnIlePheAspLysGlnLysLysTyrAspGluIleLys 401
OY 1225 TCATAT-----TTATGACAGATTAACAAATTTGTCAAT 1257
Db 402 LysTyrGluAsnGlyLysAspSerArgGlnLysArgAspAlaGlyThrThr 421
OY 1258 AATATTAAATAGGAA---TATTATAACATTTATGAAAACTTAAGAAACGCAATAT 1314
Db 422 ThrThrAspTyrAspLysGlyLysPheTyrAspGluLeuAsnLysSerGluTyr 441
OY 1315 GCAACTAATGACACTTTTAAATTAATTAAGAAAGAACTATGT----- 1362
Db 442 ArgThrValAspLysPheLeuGluLysLeuSerAsnGluIleCysThrLysValLys 461
OY 1363 -----AAAGAGATTAACAGAGAAAGATTAATTAATTAAGTGT----- 1410
Db 462 AspGluGluGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyVala 481
OY 1411 -----GATGACAAAGGATATTTATGCTTACAGAAATATGCAAGTGT 1455
Db 482 SerGlyThrAsnValGlnSerGlnLysThrPheTyrArgSerLysTyrCysGlnProCys 501
OY 1456 CCCGACGCGGGGTCGAATGTGATGATTAATTAATACACACAC-----AATCGAATTAATGAT 1512
Db 502 ProTyrCysGlyValLys-----LysValAsnAsnGlyLysSerSerAsnGln 517
OY 1513 CGTGAACGTGTAATAT-----GAAGACTAATAACCTCCATGGGT 1554
Db 518 TrpGluGluLysAsnAsnGlnLysCysLysSerGlyLysLeuTyrGluProLysProAsp 537
OY 1555 GTGAGCGCTACATATATCATCTGCTGCTTTATAGTGGATATGAACAAAGGTATTAACAA 1614
Db 538 LysGlnGlyThrThrIleThrIleLeuLysSerGlyLysGlyHisAspAspIleGluGlu 557
OY 1615 AATTAAGAAATTTTGTATAC-----AGCTCA 1641
Db 558 LysLeuAsnLysPheCysAspLysLysAsnGlyAspThrIleAsnSerGlyLysSerGly 577
OY 1642 ACTAATTAACAAAGTAAATAT-----CAAAATGGGAATGC 1680
Db 578 ThrGlyGlySerGlyGlyLysAsnSerGlyArgGlnGluLeuTyrGluGluTrpLysCys 597
OY 1681 TATTATAAGATCAAAATTAATTAATAGTAACTGAAACAAACAAATACGAATCAATATAT 1740
Db 598 ---TyrLysGlyGluAspValValLysValGlyHisAspGluAspAspGluLysAspTyr 616
OY 1741 GATTAATCCTTAAG-----ATAATA----- 1758
Db 617 GluAsnValLysAsnAlaGlyLysLeuCysIleLeuLysAsnGlnLysLysAsnLysGlu 636
OY 1759 -----TCATTTCATTAATTTT 1773
Db 637 GluGlyGlyAsnThrSerGluLysGluProAspGluIleGlnLysThrPheAsnProPhe 656
OY 1774 TTTGAATTAATGCTTACATATTTATTAAGGATACTATTAAGATGCAATGCAAACTTAA 1833
Db 657 PheTyrTyrTrpValAlaHisMetLeuLysAspSerIleHisTrpLysLysLeuGln 676
OY 1834 ACTGTATA---AATATACACACACGATGT---ATGATGAATGTAAACAGAAATGTC 1887
Db 677 ArgCysLeuGlnAsnGlyAsnArgIleLysCysGlyAsnAsnLysCysAsnAsnAspCys 696
OY 1888 TTATGTTTGAACATGCTTAAACAAAGAAAGAAAGAAATATATTAAGAAAGAACTG 1947
Db 697 GlucysPheLysArgTrpIleThrGlnLysLysAspGluTrpGlyLysIleValGlnHis 716
OY 1948 TTCACAAAAAAGAAATATACAG----- 1971
Db 717 Phe---LysThrGlnAsnIleLysGlyArgGlyLysSerAspAsnThrAlaLeuIle 735
OY 1972 -----CAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2016
Db 736 ProPheAspHisAspTyrValLeuGlnTyrAsnLeuGlnGluIlePheLeuLysGlyAsp 755
OY 2017 -----TTTAAAGTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2058
Db 756 SerGluAspAlaSerGluGluLysSerGluAsnSerLeuAspAlaGluGluAla----- 773
OY 2059 AAAGAACTTTGGAATATATTAAGAAAGAAAGAAAGAAATGAGTTTCCAAATTTGGAATTAAT 2118
Db 774 -----GluGluLeuLysHisLeuArgGluIleIleGluSerGluAspAsn 789
OY 2119 AGGAC-----TATTAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2154
Db 790 GlnGluLysSerValGlyGlyValThrGlnGluLysAsnIleMetAspLysLeuLeu 809
OY 2155 GATCACTTAAGAAAGAACTGCCAGATATGT----- 2184
Db 810 AsnTyrGluLysAspGluLysAspLeuCysLeuGluIleHisGluAspGluGluGlu 829
OY 2185 ---AAAGACATTAATTAACAAAGCAAGCATGTAACATCCATTAATGCAACAAACCG 2241
Db 830 LysGlnLysGlyAspGlyAsnGlnCysIleGluGluGlyGluAsnPheArgTyrAsnPro 849
OY 2242 TGTGTTAAACCTGTCGAGGACGCAAC---CCCACTAAATAATATTAAGAAATTAACACAA 2298
Db 850 CysSerGlyGlySerGlyAsnLysArgTyrProValAlaAsnLysValAlaIleTyrGln 869
OY 2299 TACTTTAAAGAGTGCAATACAGAGAAAGCAAGAAATCGTGTCTTCATTAATTAAGAAAGAA 2358

```

```

Db      870 MethHisHisLysLysLysThrGlnLeuAlaSerArgAlaGlyArgSerAlaLeuArgGly 889
QY      2359 AAGGCAACAGCAAGGTATATATTAAGCTGGGGGAGAGAAAGACATTCACAGCAATTTA 2418
Db      890 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlyGlnIle 909
QY      2419 TGTAGATATATGATTAACATTCATTAATCTGTTTTCATTAATGAGCAGCATGTGAT 2478
Db      910 CysLysIleAsnGlnLysSerAsnAspSerArgLysAsnSerGlyProCysThr 929
QY      2479 GCGAAAGGCAAGGTATGATATACAAACAGATTTGCTGAGCAAGTGAAGTGAAGT 2538
Db      930 GlyLys---AspLysAsnLysGlyLysAlaArgMetArgIleGlyThrGlnProSerAsn 948
QY      2539 GATCCGGAACACATGCGTAAAGATGACGAAGATGTTATGATGCTGCTGACAGACAGAT 2598
Db      949 IleGlnGlyLysLysGlnThrSerTyrLysAsnValPheLeuProProArgGlnHis 968
QY      2599 ATATGATATCCATTCATTTGGAACATTTACAAACGATGATCACCACCTTAATGTAATAT 2658
Db      969 MetCysThrSerAsnLeuGlnAsnLeuAspVal-----GlySerVal 982
QY      2659 GTT---GATGATTTATATATATTCCTTTTGGGGGAGTCTTCTATCAGCAAAATAT 2715
Db      983 ThrLysAsnAspLysAlaSerHisSerLeuGlyLysPvalGlnLeuAlaLysThr 1002
QY      2716 GAAGCAAAACAGATTAATACGATGTTAAAGAAAGATTAACCTTAAGGCCCCAAGAA 2775
Db      1003 AspAlaLalaGlnIleLysArgTyrLysAsnGlnAsnHisIle-----Gln 1018
QY      2776 GTAAGTGACCA-----AAACACAGACACATCTGTGACATATACGTTACGAT 2826
Db      1019 LeuThrAspProIleGlnGlnLysAsnGlnAlaIleMetCysArgAlaValArgTyrSer 1038
QY      2827 TTGCGATATATAGTATATTAATTCGAGAGAGATCTCGGAAAGAAAGCGTGACATG 2886
Db      1039 PheAlaAsnLysPheLysIleLeuArgLysArgAspMetTrpAspLysSerSer 1058
QY      2887 GTAAGCTGCAAGACATATGGAACATGTTTGTGATATATACATATACATCAACGCA 2946
Db      1059 ThrAspMetGlnThrArgLeuIleThrValPheLysAsnIleLysGlnLysHisAspGly 1078
QY      2947 ---AAAGAAATGATTAATATATGATGAT-----GCCCCCAATATTTAAATG 2994
Db      1079 IleLysAsnAsnProLysTyrThrGlyAsnGlnSerLysLysProAlaTyrLysLeu 1098
QY      2995 AGGGAATATGTTGGAGCTATATAGCCAAAGTATGGGAGCAAGTGAATGTGATATA 3054
Db      1099 ArgAlaAsnTrpTrpLysAlaAsnArgHisGlnValTrpArgAlaMetLysCys----- 1116
QY      3055 AATATATTTGAAGATTAATCGGACACACATCAACAAAGTATGTTATTCGGATATAGT 3114
Db      1117 -----AlaThrLysGlyLysIleLys----- 1123
QY      3115 GATCATATACCATGATGATTTATATCCCAAAATTAAGATGATGACCAATGGGCA 3174
Db      1124 ProGlyMetProValAspAspTyrIleProGlnArgLeuArgTrpMetThrGlnTrpAla 1143
QY      3175 GAATGATACGCAAGGTGCGAAAGAGCTATGATATAGTGAAGAGCAAGTGAAGAG 3234
Db      1144 GlnTrpTyrCysLysAlaGlnSerGlnGlnTyrAspLysLeuLysIleCysAlaAsp 1163
QY      3235 TGTAAAGATAGATATGCTCAAGCTGTGACAAAGAGTGCTGACAGTGTACAGAG 3294
Db      1164 CysMetSerLysGlyAspGlyLys---CysThr---GlnLysAspValAspCysGlyLys 1181
QY      3295 TGCACAGAACCTTGATATGATATATGATATATAGATATGAGAAAGCAATGGAAT 3354
Db      1182 CysLysAlaLysCysAspLysTyrLysGlnGlnIleGlnLysTrpAsnGlnGlnTrpArg 1201
QY      3355 ATATATACATTAATTAACAAGATTAACAGCAAGCAACAAATGCTGTATATATAGT 3414

```

```

Db      1202 LysIleSerAspLysTyrAsnLeuLeuTyrLeuGlnAlaLysThrThrSerThrAsnPro 1221
QY      3415 GGTATGAAAGCTTCCAGTCTCGCCAAATATCATATAGAC-----AGCAATGTATAT 3465
Db      1222 Gly-----ArgThrValLeuGlyAspAspAspProAspTyrGlnGlnMetVal 1237
QY      3466 GAATTTTGTCCGATTAATACCAACAAATGTTGGCAAA-----AGTAAATAA 3513
Db      1238 AspPheLeuThrProIleHisLysAlaSerIleAlaLysValLeuValLysArgAla 1257
QY      3514 AGTGGTACTAGTATGATAAAGTCTGTCATTCGTATACACACACGATGATAAATGTTGA 3573
Db      1258 AlaGlySerProThrGlnIleAlaAlaAlaProIleThrProTyrSerThrAlaAla 1277
QY      3574 GCATATCTCCATGATACAGAAATTTGATGATGATGTCAGTCCAAATGAGTTTGTGAT 3633
Db      1278 GlyTyrIleHisGlnGlnIleGlyTyrGlyGlyCysGlnGlnIleThrGlnPheCysGly 1297
QY      3634 GAAAA-----AGTATGTTAGGATTAACGAAATATGCTTTAGA 3675
Db      1298 LysLysHisGlyAlaThrSerThrSerThrThrLysGlnLysGlnLysGlnLysThrPheLys 1317
QY      3676 GATTAACACAGACCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3735
Db      1318 GlnProProGlnIleTyrAlaThrAlaCysAspCysIleAsnArgSerGlnThrGln--- 1336
QY      3736 GTACAGATTAATAAAGCAAAAGCAAGAAAGCAAGATGCAAGATTTGCAAAACAGTGAAT 3795
Db      1337 -----GlnProLysLysLysGlnGlnLysAsnValGlnSerLysLysIleValGln 1353
QY      3796 GATATACTTAAGCAAAAGCAAGTGAAGAAAGCAAGTGAAGATTTGCAAAACAGTGAAT 3855
Db      1354 LysIleLeuGlnGlnLysAsnGlyArgThrThrValGlyLysAsnProLysGlnSer 1373
QY      3856 AGTAAAGATTAACCCGATGCGCAATGC-----GCAATTAATATATGCAAGACCTGCT 3912
Db      1374 -----TyrProAspTrpAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1389
QY      3913 GGTGTATGCCCCCTGAGAGACAAAGTATGCGTATGTTGTCGAATGATATGA 3972
Db      1390 AlaCysMetProProArgTrpGlnLysLeuCysLeuTyrTyrIleAlaHisGlnSerGln 1409
QY      3973 ATTAATAATTAATCAATCAACAGTATTAATAAAGAGCTTTCATCAATCTGCACAGCA 4032
Db      1410 ThrGlnAsnIleLysThrAspAsnLeuLysAspAlaPheIleLysThrAlaAlaAla 1429
QY      4033 GAAACATTTCTTCATGATGATTTATTAAGTAAAGTAAAC---GATGCGTGAAGAAATGAATC 4089
Db      1430 GlnThrPheLeuSerTrpGlnTyrTyrLysSerLysAsnAspSerGlnAlaLysIleLeu 1449
QY      4090 GATTAAGATTAATAAAGAGCAAAATTCCTCCGCAATTTTGGATCCATGTTCTACACA 4149
Db      1450 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetMetTyrThr 1465
QY      4150 TTGCGATATATAGATATTTTATTTGGAACAGATATATCAAAAGCTCATGGTGAAGCA 4209
Db      1466 PheGlyAspTyrTrpArgLysPheLysLeuAsnThrAspIleSerLysGlnAsnAspVal 1485
QY      4210 AGTAAACTAAAGACGCAATAGATTCCTCTTTGCAAAAGGTGACCAAAATTCCTGAT 4269
Db      1486 AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspLysSerLysSerProSer 1505
QY      4270 GCAAAACACAGCCCAAGAAATGTTGACAGACAAATGATGATATGAGCAAGTATGCA 4329
Db      1506 GlyLeuSerArgGlnGlnIleTrpTrpLysThrAsnGlyProGlnIleTrpLysGlyMetLeu 1525
QY      4330 TGTGCACTAGTAAATTTGGGCAAAAGATGAT-----TTTACCGAAACTAC 4380
Db      1526 CysAlaLeuThrTyrValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr 1545
QY      4381 GGTCAACAACAGCTCAATTTACTGACAAA---AGACACACCTTGAGGAATTTGGCAAA 4437
Db      1546 SerTyrAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnGlnPheAlaAla 1565

```

```
QY 4438 CGACCCAGTCTTTTACGATGGCTTAACCGAATGGTACGACATTTGCTATACAGCA 4497
Db 1566 LysProGlnPheLeuArgTyrMetIleGluTyrPheGluLeuPheCysAlaGluThrGln 1585
QY 4498 AATATATTGAAGATGTGCAGGAAAAATGTAACTCA--AATGACCAATTTGAAGTGAT 4554
Db 1586 LysLysGluAsnIleLeuLysAspAlaCysAsnGlnLeuLeuSerThrGlnGlnCysAsn 1605
QY 4555 ACA-----GAATGATATAGAAATGGAGACACTTAAATATATGAAA---AAA 4602
Db 1606 AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGluTyrAlaGluAsnLys 1625
QY 4603 AAAAAAGATGGATTCACACAGATTAATATATTACAGGATGACGCAAAAAAAGATTC 4662
Db 1626 LysLysGluPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln 1645
QY 4663 GATAGACAAACATGTGTGTAAATGCTTACAGACTATACGAAAGAAATGCACAGATTAC 4722
Db 1646 AspProGluTyrLysGlyTyrGluTyrLysAsp----- 1656
QY 4723 TTGACAGAGAAATTTACTGCTAGTGTGATAGCTGGAAGTGCCTGTGTGATCAA 4782
Db 1657 -----GlyValGlnPro-----IleGln 1662
QY 4783 AGAATATATCAATTTGTAGAAAAACAGGCTTACTATGATCCGACAAACATTTGTGGTGC 4842
Db 1663 GlyAsnGluTyrLeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1678
QY 4843 ACAAAATTTATGTGAATATGACGCAAAATATACTAATTTTCAGTAAATATAGTCCAAA 4902
Db 1678 ----- 1678
QY 4903 GCATTAGTAAAGAGGCAAAACACAGGTGCTATTAACTGCAAAACAAAGTCTATATAC 4962
Db 1678 ----- 1678
QY 4963 TACAATTAATGAAAGATTTGACTGAAGATGTGCTTTTCTTCTTCGCTGACTACTATA 5022
Db 1678 ----- 1678
QY 5023 TGTTCATGCAATTTGATGCAATTTATACAGATCCAGAGTTAAAGATGAAATGGGTTG 5082
Db 1679 -----MetAspLysAsnValLeuSerValSerProLysGluLys----- 1691
QY 5083 CGAAAAAGATTGATGGAAGTGGCGCAACGGAAGGTTACATTTGGTCAATACTACAAA 5142
Db 1692 -----ProPheGlyLysTyr----- 1696
QY 5143 GAAAAAAGAAAAAGAGCAAAATTAATAACGTCGATGGCGCAAAATTTCTTATGAGTTC 5202
Db 1697 -----AlaHisLysTyrPro----- 1701
QY 5203 CCGCCTTGTAGTGTATGAATATAGTTTATGATTAAGAGATATATCTAGTATTT 5262
Db 1701 ----- 1701
QY 5263 GATATTTGGAAGATGAAAAACAAAAGACCGAAGAAATTTGAGAAAAATATTTACAAA 5322
Db 1701 ----- 1701
QY 5323 AATGGAACATCAGTTGGCAAAAGAGTATAGTACTACAGAAATCCGGTAGTACTGCG 5382
Db 1701 ----- 1701
QY 5383 CGAAAAATTTTCTGGAACGAAAAATAGAANGTGTGTGAACGCAATGATATGCGGCTAC 5442
Db 1701 ----- 1701
QY 5443 AAACGTGTAGGATGATGAAATAGTGAATATAGTGAAGAAGTATGAAGATCTTAAA 5502
Db 1701 ----- 1701

QY 5503 AATGTGTTCTGTACTTTCACATGATGATTAATCTTANGGGAAAAATCCGATGAAGT 5562
Db 1701 ----- 1701
QY 5563 ACTGCGTATCAGTTTCTTCGATGATGTTTGGCGAATGGGGTGAAGATTTTGCAAACATAAA 5622
Db 1701 ----- 1701
QY 5623 GAAAAAGAAATTTGAGAAATTTGGTAGGGCGGTATATGATTAATCTTGTGTATATGAA 5682
Db 1701 ----- 1701
QY 5683 GATTAAGAAAAATGTACAGATCCGTACACATATTAATAAATTTATTAAGTACTGG 5742
Db 1702 -----GluLysCys----- 1704
QY 5743 AAACACAGTATGAAAAACAAATCAAAAAATATGTGTAGAAATTAAGCAAAATATATTC 5802
Db 1704 ----- 1704
QY 5803 GAGCATCTGTGGCAAAAGATGACAGAGAGCGCTGCGAATATTAACAACAAATTA 5862
Db 1704 ----- 1704
QY 5863 AAAATTTGTGAATAAATAAGTGAAGATTTGATGATATAGTATGAAGAATGTGTCACA 5922
Db 1705 -----AspCys----- 1706
QY 5923 CAGCATTAATGATGTTAATAGTCAAAAAATATGCCCCGATCATTAAGATGAACCAAAA 5982
Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAGTGAATTTGCAAGTGGCCAGAGTCCACAGCTGACAGATGACAGAA 6042
Db 1707 -----TyrGlnGlyLys-----HisValProSerIleProPro----- 1717
QY 6043 ACACGTCACACGCGGTATCACTGATATCAAAACGACGCGCATGCAAAAAAGAACGAAA 6102
Db 1717 ----- 1717
QY 6103 ACAGCGCGCCTACAAAAACAGCCGAAAAAGTGAATACTAACACAGAAATGCGAGCA 6162
Db 1718 -----ProProProProValGlnProGln----- 1725
QY 6163 CAACACGAAACCCGACGACGACACAAACAAACAGAAAGAACATCAACAGCAACA 6222
Db 1726 -----ProGluAlaProThr----- 1730
QY 6223 ACAGATCTGACGTGGCACACANGCTAAAGCCATTTCTTGATAAACCAGATAGCAGG 6282
Db 1731 ValThrValAspValCysSerIleValLysThrLeuPheLys-----AspThrAsn 1747
QY 6283 GGTGAATAGAGGGTCTAATCCAAAAGCATATGACAAAT-----CCTAAATGGGT 6336
Db 1748 AsnPheSerAspAlaCysGlyLeuLys--TyrGlyLysThrAlaProSerSerTyrLys 1766
QY 6337 TGTATTT-----GTAGTAAGTCAATAAGAAAT 6363
Db 1767 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrGlyLysSerGlySerAsp 1786
QY 6364 GAAATGCAATATGATAGCCTCTAGAGAGAAAAAATTTATGTAATAATATATACATAT 6423
Db 1787 SerGlySerIleCysIleProArgArgArgArgLeuTyrValGlyLysLeuGln--- 1805
QY 6424 TTAATATGAACT-----GAAATATAGCTGAC 6453
Db 1806 ---GluTyrPalaThrAlaLeuProGlnGlyGluGlyAlaAlaProSerHisSerArgAla 1824
QY 6454 AATGATTAAGAAGGCTTTTATTAATGTGACAGATAGAAATCAATTTTGTGTTA 6513
Db 1825 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuTyrAsp 1844
QY 6514 AATATATATTAATTGAA---AATCTGACAGCAAAAAATGAA----- 6549
```

Db 1845 ArgTyrLysGluGlnLysLysProGlnGlySerGlnGlnAlaLeuSerGlnLeu 1864
QY 6550 -----TTCAGAAAATGGAACA 6564
Db 1865 ThrSerThrTyrSerAspAspGluGlnAspProProAspLysLeuGlnAsnGlnLys 1884
QY 6565 ATTCACAGATCAATTTAAAGCAATATCTATTATACATATGCTGATTTATAAGATATGTTT 6624
Db 1885 IleProProAspPheLeuAlaGluMetPheTyrThrLeuGlnAspTyrArgAspIleLeu 1904
QY 6625 TTTCGAACGTATATTTCTAATGATATAAAATATATACCTGATACCAATAGCTATACACACC 6684
Db 1905 ValHisGlyLysGlnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1924
QY 6685 ATTCCTCAATGAAATATATAAGAAAACAGATAAAAAAAAGTAAAGATTAACGTAA 6744
Db 1925 ValLeuGlnAlaSerGlyAsnLysGlnAspMetGlnLysIleGlnGlnLysIleGlnGln 1944
QY 6745 ATA----- 6747
Db 1945 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1964
QY 6748 ----TTTGGAGAGAAAATTAATAATTTATTTGGAGAGATGATATGATTAACCTTAT 6804
Db 1965 LysTrpTrpAsnGlnHisAlaGlnSerIleTrpLysGlyMetIleCysAlaLeuThrTyr 1984
QY 6805 -----CATCTCAGACAGAAAACGAAAAAGAAAAAATAATAGA 6840
Db 1985 ThrGluLysAsnProAspThrSerAlaArgGlyAspGlnAsnLysIleGlnLysAspAsp 2004
QY 6841 GATATATAC----- 6849
Db 2005 GluValTyrGlnLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2024
QY 6850 ----CAGTACAAAT--GACATGACCAACG----- 6873
Db 2025 ThrGlyThrTyrLysThrGlnTyrAspTyrGlnLysValLysLeuGlnAspThrSerGly 2044
QY 6874 -----ACGCTTCCTCCCTGAGAGATTTGTAATAAGGGCC 6906
Db 2045 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheValIleuArgPro 2064
QY 6907 CAATTTTGGAGATGCTTACAGAGATGGCAGAGAAATTTGTAATAAGAGAGAGAAACAG 6966
Db 2065 ProTyrPheArgTyrLeuGlnGlnTrpGlyGlnAsnPheCysLysAspGlyHisLys 2084
QY 6967 TTGTTAAATTTGAGCGCGGCTGTAAGCAATATGAG----- 7002
Db 2085 LeuAlaGlnIleLysHisGlnCysLysValGlnGlnLysnGlnGlyLysSerArgArgGly 2104
QY 7003 -----TGTAAAT-----GGTAGTAAT 7017
Db 2105 GlyIleThrArgGlnTyrSerGlyAspGlyGlnAlaCysAsnGlnMetLeuProLysAsn 2124
QY 7018 GAGCGTGAACACAGAA-----TGTCCAGAGCGGCTGTGACATATCAAA 7062
Db 2125 AspIleThrValProAspLeuGlnLysProSerCysAlaLysProCysSerSerTyrArg 2144
QY 7063 AATTTTAAAGAAAGTGAAGAACTGATATGAACAGAAAGAAAGTCAAAAGAT 7122
Db 2145 LysTrpIleGlnSerLysGlyLysGlnPheGlnLysGlnGlnLysAlaTyrGlnGlnGln 2164
QY 7123 AAGAT----- 7128
Db 2165 LysAspLysCysValAsnGlnLysSerAsnLysHisAspAsnGlnLysCysGlnThrLeuThr 2184
QY 7129 ----GGCAAAAAGTAAAGATAT-----CCTTCTACTGAAAGACATA 7170
Db 2185 ThrSerGlyLysAlaLysAspPheLeuLysThrLeuGlnLysProCysLysProAsnVal 2204
QY 7171 GAGAAAGCAACATGCTGCATGATATTTAAACATGAATTAAGAAATATATGCGCAAT 7230

Db 2205 GluGlyLysThrIlePheAspAspAspLysThrPheLys-----HisThr 2219
QY 7231 AAGATTTGTTCTGTATGCAAAAACCTTCTCACAACTACCAAAAACACCAACAATCA 7290
Db 2220 LysAspCysAspProCysLeuLysPheSerValAsnCysLysLys----- 2234
QY 7291 CAATCATCCATGCTAATGATATGATCCAGAAATCGCTGATATATGTTCTCTCAAGAAATTTAAC 7350
Db 2235 -----Asp 2235
QY 7351 AAGTGAGTGCTCCTGAACTTTCAAAAAGGAGCTATGATCATCAAAAAAATTAAT 7410
Db 2236 GluCysAsp-----AsnSerLysGlyThrAspCysArgAsnLys----- 2248
QY 7411 GAACCTAAATTAACCTATGAAATGTGTAGAGAAAGACATATTTATTTATTAAGAGCA 7470
Db 2249 -----AsnSerIleAsp-----AlaThrAspIle 2256
QY 7471 GAAATATATATGATATATACCTTGAAGAGAAAATTTATACCTATAGCTTACAAAGAA 7530
Db 2257 GluAsnGlyValAspSerThrVal-----LeuGlnMetArgValSer 2270
QY 7531 AAGGAAAGTAAATAATGTTGACATAATATATATCCTGCGATCCTAAGAAACCTTATGCA 7590
Db 2271 AlaAspSerLysSerGlyPheAsnGlyAsp----- 2280
QY 7591 CCGATTAATATATATAGAGAAAGAAACCTTGTAAT----- 7629
Db 2281 -----GlyLeuGlnAsnAlaCysArgGlyAlaGlyIlePheGlnGlyIle 2295
QY 7630 AGAGAGAAATATGCTTTTAAGTAGATTAATGATGAAATGTTACAAAATTCAAAGTTC 7689
Db 2296 ArgLysAsp-----GluTrpLysCys--ArgAsnValLysGly 2307
QY 7690 TATCAGAGAGAAAAAGATGATGTACCTCAAGAGAGAACATATGCTCTTAAGGAAT 7749
Db 2308 TyrVal-----ValCysLysPro----- 2313
QY 7750 TTAGATGAATTAATAATTTGAAGACTTGAAGATGATTAATATCTCTTAATAATGTTTCGT 7809
Db 2313 ----- 2313
QY 7810 CGAACTCAGCAAAATGAAGCAATAGACATATATATAAAACCTTCAACTCAGAAACGGTGC 7869
Db 2314 -----GluAsnValAsnGlnGlu----- 2319
QY 7870 GCAATGAATCCAAATATGTATGATATGAAATATAGTTTCCTGATCGGTGACATAGT 7929
Db 2320 -----Ala 2320
QY 7930 AGAGAAACAGATATGTTACGAATTTGTTACTTACCTCCGTAAGAAATTAATATAT 7989
Db 2321 LysGlyLysHisIleIleGlnIleArgAlaLeuValLys-----ArgTrpVal 2336
QY 7990 AAGTTTGAATATATATATGAAAATGGAAGAAATATAAATTAAGSTACAAATTAATAC 8049
Db 2337 GluTyrPhePheGlnAspTyrAsnLysIleLysHisLysIleSerHisArgGlyLysAsn 2356
QY 8050 AAGCAT-----GATCAAAAGCTTTCCTGCTGCTGGGAGTGTATATAGAAA 8097
Db 2357 GlyGlnLysSerProCysIleLysAsnCysValGlnLysTrpValAspGlnLysArgGln 2376
QY 8098 GATATTTGAAGCAATAGCTGCAAGACACAGAAAGATCAAAACTTTTAGAAAAGGA 8157
Db 2377 Glu--TrpLysGlnIleThrGlnArgPheLysAsp----- 2387
QY 8158 AGAATGATGATTTGAACGCAATACATTTAATACAGATAGCTGACATTAAGGAGAT 8217
Db 2388 -----GlnTyrLysAsnAspAsnSerAspAsp 2396
QY 8218 CAACCTGTGATGATATATATACCTCAACGGTTTCGATGATGATGATGATGATAT 8277
Db 2397 AspAsnValArgSerPheLeuGlnThrLeuIleProGlnIleThrAspAlaAsn----- 2414


```
D 232 GluGlnLysLeuLysAspIlePheLysIleHis-----LysAspValMetLys 248
QY 733 AATGAT-----TACAACTCGATGGA---TCGTGAATATTATTAATTA 774
D 249 ThrAsnGlyAlaGlnGluArgTyrIleAspAspAlaLysGlyGlyAspPhePheGlnLeu 268
QY 775 AGAAGACATGCGGAATGGAATAGAAATAGATGGAACCTATACATGATGCA 834
D 269 ArgGluAspTrpTrpTrpSerAsnArgGluThrValTrpLysAlaLeuIleCysHisAla 288
QY 835 TCATATAATCTGATATTATTATGCAATGAAAGTAATACCATATTATTCAATCT 894
D 289 ProLysGlnAlaAsnTrpPheIleLysThrAlaCysAsnValGlyLysGlyThrAsnGly 308
QY 895 AATGCGGCATTAACAGAGAAAGGTCCTACCAATTTAGTATGTCCTCATATTTA 954
D 309 GlnCysHisLysGlyIleGlyLysAspValProThrTrpPheAspTyrValProGlnTrpLeu 328
QY 955 CGTTGCTTCAGACATGCGGAGAGAAAGCTTTGCCGAAAAAATAATTAATTGA 1014
D 329 ArgTrpPheGlnGluTrpAlaGlnAspPheCysArgLysLysLysLysLeuGlnAsn 348
QY 1015 GTCAGAGACCTCGTGTATGACAAAGACGTTATATGTAGTCTATATGACATGAT 1074
D 349 LeuGlnLysGlnCysArgAspTyrGlnGlnAsnLeuTrpCysSerGlyAsnGlyTyrAsp 368
QY 1075 TGTACGACCACTATTGGAAGAAAGATTTGTCATTTGATATATAAGTACTGAT 1134
D 369 CysThrLysThrIleTyrLysGlyLysLeuValIleGlyGlnHisCysThrAsnCys 388
QY 1135 TCCACATTAATGCAAGCTTTTGAAGTTGCTTGGGCAATACACAGCATTTAAAA 1194
D 389 SerValTrpCysArgMetLysTrpGluThrTrpIleAspAsnGlnLysGlnPheLeuLys 408
QY 1195 CAAAAAGAAAAATATGAAAAAATA-----1221
D 409 GlnLysArgLysTyrGlnTrpGlnIleSerGlyGlyLysSerProLysArg 428
QY 1222 -----CAATCATATTATGCAAGCATACAAATTTGCAATATATT 1263
D 429 ThrLysArgAlaAlaArgSerSerSerSerSerAspAsn-----442
QY 1264 AATAGTCATATTATTAACAATTTTATGAAAACTTAAAGAAAGCATATGCACTAAT 1323
D 443 -----GlyTyrGlnSerLysPheTyrLysLysLeuLysGlnValGlyTyrGlnAspVal 460
QY 1324 GACACTTTTAAATTACTTAATGAAAGAAAGATTTGTAAGA-----GGA 1371
D 461 AspLysPheLeuLysIleLeuAsnLysGlnGlyIleCysGlnLysGlnProGlnValGly 480
QY 1372 TTACCAAGAGAAAAAGATATTACTTTTACTTAACAGTCTGATGACAAAGGATATTAT 1431
D 481 AsnGlnLysAlaAspAsnValAspPheThrAsnGlnLysTyrValLys---ThrPheSer 499
QY 1432 CGTTACGAAATATTGCCAAGTGTCTCCGACTCGGGGCTCAATGATGATTAATAATAC 1491
D 500 ArgTrpGlnIleCysGlnProCysProTrpCysGlyLeuGlnLysGlyLysProProTrp 519
QY 1492 ACACACCAATCATATATGATCGTAACGTGTAAATATGACAGCATTAACCTCATG 1551
D 520 LysValLysGlyLysPheLysThrCysGlySerAlaLysThrLysThrTyrAspProLys---538
QY 1552 GGTGTGAGGCTACTAATATCACTGCTCTTATAGTGTATGACAAAGGATATTACA 1611
D 539 -----AsnIleThrAspIleProValIleuLysTrpAspLysSerGlnAsnIleLeu 556
QY 1612 CAAAAATTAATAATTTTGTAAACAGCTCAACTAATTAACAAGATTAATAATCAAAA 1671
D 557 LysLysTyrTrpLysAsnPheCysGlnLysAlaPro---GlyGlyGlnIleLysLys 575
QY 1672 TGGGATGCTATTATATGAT-----GAAATATTAATATAGATGAACCTG 1716
D 576 TrpGlnCysTyrTyrAspGlnHisArgProSerSerLysAsnAsnAsnCysValGln 595
QY 1717 GAACAAAATGACTGAATCATTAATGATTAATCTTAAGATTAATCATTTTCAATTTT 1776
D 596 GlyThrTrpAspLysPheThrGlnGlnLysGlnThrValLysSerTyrAsnValPhePhe 615
QY 1777 GAATTAAGGCTTACATATTATTATTAAGGATACATTATTAAGTGAATGACAACTTAA 1836
D 616 TrpAspTrpValHisAspMetLeuHisAspSerValGlnTrpLysThrGlnLeuSerLys 635
QY 1837 TGTATAATTAATCAACCCAG-----CATTTATGATGAATGTAACAGAAAT 1884
D 636 CysIleAsnAsnAsnThrAsnGlnLysAsnThrCysArgAsnAsnAsnLysCysLysThrAsp 655
QY 1885 TCGTTATGTTTGTACAGATGCGTTAAACAAAAGAGAAAGAAAGATGATATTAAGAA 1944
D 656 CysGlyCysPheGlnLysTrpValGlnLysLysGlnGlnGlnTrpMetAlaIleLysAsp 675
QY 1945 CTGTTCACAAAAAAGAAATATACAGCAATCGTATTATTAATTAATTAATTTT 2004
D 676 HisPheGlyLysGlnThrAspIleValGlnGln-----LysGlyLeuIle 690
QY 2005 GAAGTTATTTTAAAGTATGATTAAGTAACTGACAAAGATGAACAAATGGAAGAA 2064
D 691 ValPheSerProTyrGlnValLeuAspLeuVal-----LeuLysGlyLysAsn 706
QY 2065 CTATGCAAAATATTAATAAGAAAAAATAAGCTTTTCCAAATTTGGAATAATATAGGAC 2124
D 707 LeuGlnGlnAsnLysLysAspValHisGlyAspThrAspAspIleLysHisIleLysLys 726
QY 2125 TATTTAGAC-----AATGCAATA 2142
D 727 LeuLeuAspGlnGlnLysAspAlaValAlaValLeuGlyLysLysAspAsnThrTrp 746
QY 2143 GAACCTGTGTAGTACCTTAATAAGAAACTGCCAGCATATTAAGCAATATATACAAAC 2202
D 747 AspLysLeuLeuGlnHisGlnLysGlnGlnAlaGlnGlnCysLysGlnLys-----Gln 764
QY 2203 GAAGCATGTGCAACATCCCATTAATGCAACAACAACCCGTGTGTTAACTCGTGAGGC 2262
D 765 GlnGlnLysGln-----768
QY 2263 AGCCAAACCCACTTAATAATTAATAAGAAATAGCAATACTTTAAAGAGCTGATCGAG 2322
D 769 -----LysLysAlaGlnGln-----773
QY 2323 GAAGCAGAAATGCTGTCTTAAATTAATGAAGAAAGCAGCAGAGATATATATAA 2382
D 774 GlnSerArgLysArgSerAlaGln-----781
QY 2383 CGTGGGGGTAGAGAAAGACTTCAAGACAAATTTATGTAGATTAATGAATAAACATTC 2442
D 782 -----ThrArgLysAspGlnArg-----787
QY 2443 AATCGTATCTGCTTTTCAATGACACATGATGCGCAAGCAGCAGGATGCTATA 2502
D 788 -----ThrGlnGlnProAlaAspSerAlaGln-----797
QY 2503 CAAACAAGATTGTCTAGAACCTGAATGGAAGTGGATCGGAGACATGCTTAAGAT 2562
D 798 -----ValGlnGlnGlnGlnAspAspAspAspAspAspAspAspAspAspAsp 810
QY 2563 CACGAAGAGCTTATATGCTCTCTAGAAAGCAGCATATATGATCACTTTTGAAACAT 2622
D 811 AspGlnAsp-----813
QY 2623 TTACAAAGCATGATACCCACTTAATGTAATATTGTTGATGATTTAGTTAATTC 2682
D 814 -----AspAspValAlaGlnGlnGln-----820
QY 2683 TTTTGGGGGATGTCTCTATACAGCAAAATATGAAGCAAAAGATATAGCAATGTAT 2742
D 821 GlnGlnGln-----823
```

QY 2743 AAAGAAAGATTAACCTTAAGGCCCCCAAGAAAGTAAGTACCCAAAACAGCAACT 2802
 |||||::: |||||::: |||||::: |||||:::
 Db 824 LysGIuGIuGIuThrValThrGIuValThrGIuValThrGIuValThrGIuValThr 842
 QY 2803 ATCTGTGACGCTATACGCTTACAGTTTCCACATATAGGTGATATATATTCGAGAGAAGAT 2862
 Db 842 ----- 842
 QY 2863 CTCTGGGAAGAAACGCTGACATGTGTAAGCTGCAGACATTTGGAACCTTTTGGT 2922
 Db 842 ----- 842
 QY 2923 AATATACATAGTCACTCAAAAGCCAAAGAAATGATATATATATGATGATGCCCCAAA 2982
 Db 842 ----- 842
 QY 2983 TATTTAAATTTGAGGAAATTTGTGGGAAGCTAATAGAGCAAAAGTATGGAAAGCCATG 3042
 Db 842 ----- 842
 QY 3043 AAATGTGATATAAATATTTGGAAGATAAATCGGGACCAATCAACACAAGATGTTAT 3102
 Db 842 ----- 842
 QY 3103 TCGGATATAGTATCATACACATTTGATGATTTATCCCAAAAATTAAGATGGATG 3162
 Db 843 ----- 843
 QY 3163 ACCGAATGGCAGAAATGTACTGCAAGGTGCAGAAAAAGAGTATGATTAAGTGAAGAG 3222
 |||||::: |||||::: |||||::: |||||:::
 Db 844 ThrGIu-----GluGIu 847
 QY 3223 AAGTGTAGAGAGTGTAAAGATTAAGATATAGTCAAGGCTGTACGAAAGAGAGTGTACA 3282
 |||||::: |||||::: |||||::: |||||:::
 Db 848 GlyValLysProCys----- 852
 QY 3283 GGTGTACGAAGTGCACAGAGCTTGTATGATTAATGATTAATGATTAATGATTAATGATA 3342
 |||||::: |||||::: |||||::: |||||:::
 Db 853 ----- 858
 QY 3343 GAACATGGAAT 3402
 Db 858 ----- 858
 QY 3403 GTTAGTAATAGTGTATTTAGAGCTTCCAGTACTGCCAAAAATCATATATACAGCAAGATGTT 3462
 Db 858 ----- 858
 QY 3463 ATTGAATTTTGTGGAATTAATACCAACAATAATGTGCGCAAAAGTATATAAGTGTACT 3522
 |||||::: |||||::: |||||::: |||||:::
 Db 858 ----- 858
 QY 3523 AGTGATGAAGTGTCTCATTTGCTACTTAACACCAGCTATGAAAAATGTGGACATATCTC 3582
 Db 858 ----- 858
 QY 3583 CATGATACAGAAATTTGATGATTTGTCAGTCAACAATAATGATTTTGTGATGAATAAACT 3642
 |||||::: |||||::: |||||::: |||||:::
 Db 859 -----PheGIuAspLysSerLeuLysGIu----- 867
 QY 3643 GATGTAGGATACGAAAAATATATGCTTTAGAGATTAACACAGAGACCATGATGTGCG 3702
 |||||::: |||||::: |||||::: |||||:::
 Db 868 -----Ala 868
 QY 3703 TGTGGTTTAAAGTGTATGATCGAAACCGACAGAGGTACAGATTAATAAAGAAAAAAGCG 3762
 |||||::: |||||::: |||||::: |||||:::
 Db 869 CysGIuLeuLysTyrGIuPro----- 875
 QY 3763 GAAGAAAGATACGGAATGTAAACAGTGAATGATTAAGTAAAGAAAGAGAGTGAAG 3822
 |||||::: |||||::: |||||::: |||||:::
 Db 876 -----GlyGIuLys 878
 QY 3823 AAACAAGTATGAAGATTTGCATCCAAAAAAGATTAAGTATCCGATTTGGCAATGC 3882

Db 879 GluLys-----PheProAsnTrpLysCys 886
 QY 3883 GGAATATATAATTTATG-----GAAGACCTCGTGTGATATGCC 3924
 |||||::: |||||::: |||||::: |||||:::
 Db 887 ValThrProSerGIuValSerThrAlaThrSerGIuLysAspGIuAlaIleCysValPro 906
 QY 3925 CCTAGAAAGACAAAAGTTATGCTTACATTTCTTGCGA-----AATGATAT 3969
 |||||::: |||||::: |||||::: |||||:::
 Db 907 ProArgArgArgLeuLysValGIuGIuLysSerGIuTrpAlaSerArgGIuLys 926
 QY 3970 GAATATAAATAATTAACATACAGAT-----AAT 3999
 |||||::: |||||::: |||||::: |||||:::
 Db 927 GluThrThrGIuValSerSerGIuAlaThrSerAlaProSerGIuSerGIuLys 946
 QY 4000 TTTAAAGAACTTTCATCAATCTGCAGCAGCAAGAACTTCTCATGATATATAT 4059
 |||||::: |||||::: |||||::: |||||:::
 Db 947 LeuArgThrAlaPheIleGIuSerAlaIleGIuThrPhePheLeuTrpHisLysTyr 966
 QY 4060 AAA-----AGTAAGATGCT----- 4074
 |||||::: |||||::: |||||::: |||||:::
 Db 967 LysGIuGIuLysLysProProAlaThrGlnAspGIuAlaGIuLysGIuValSerLeuPro 986
 QY 4075 -----GAAGAAATGAACTCGATAAGATTAAGAA-----GGCAAAATCTCT 4119
 |||||::: |||||::: |||||::: |||||:::
 Db 987 GluProSerProProGIuLysProGIuLysProGIuLysGIuLysGIuLysGIuLys 1006
 QY 4120 CCCGATTTTTCAGATCCATGTTCTTACACATTTGAGATTTATAGATTTTATTTGA 4179
 |||||::: |||||::: |||||::: |||||:::
 Db 1007 ProAspPheLeuArgGlnMetPheTyrThrLeuAlaAspTyrLysAspIleLeu 1024
 QY 4180 ACAGATATATCAAAAGTGTATGAGGAGTAACTTAAGACCAATATGATTTCTT 4239
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4240 TTCAAAAATGTTGACCAAAAATCTCTTAATGAAAAACCGCAAGATGTGCAGAGAA 4299
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4300 CATAGTATGAGATATGGGAAGCTATGCTATGTGACATAGTAAATTTGGGCAAAAAA 4359
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4360 GATGATTTTACCGAAACTACGATTACACACGTCAAATTTAGTACAAAAGACCACT 4419
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4420 TTGAGGAATTTGCCAAAGCAAGCCAGTTTACGATGCTAACCGAATGTGACGAGAC 4479
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4480 TATTGCTATACACGCAAAAATATTTGAGAGATGTGCAGAAAAATGATCAATGAC 4539
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4540 CAATTGAAGTGTGATACAGAAATGTAATGAAGAAATGCGAGACTACGTTAATATATGAAA 4599
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4600 AAAAAAAGAGTGTATTCACAAAGATTAATTTACAGAGTACAGCCGCAAAAAAAGA 4659
 |||||::: |||||::: |||||::: |||||:::
 Db 1024 ----- 1024
 QY 4660 TTCATAGACAACATTTGTTATGTTATGATACAGACTATACGAAAGATGCAACAGAT 4719
 |||||::: |||||::: |||||::: |||||:::
 Db 1025 -----TyrSerGIuSerAsnAspTrpSer 1032
 QY 4720 TACTTGACAGAAATTTACTGCTAGTTGTGTGATTAAGCTGGAAGTGTCTGTGTGA 4779
 |||||::: |||||::: |||||::: |||||:::
 Db 1032 ----- 1032
 QY 4780 CAAGAATAATATACAAATTTAGAAAAAAGAGCTTACTATGATGCCGCAAACTTGTGG 4839

Db 1032 ----- 1032
QY 4840 TGCACAAATTTATGTAATGACGACCAATATACATACATTCGACTAAGATTAAGTGC 4899
Db 1033 -----AspThrThrGlyGlyGlnThrProSerSerSerAsnAspAsnLeu 1047
QY 4900 AAGGATTAGTAAAGGAGCAACACAGGTGCTATTAAAGTGGCAAAACAAAGGCTCTAAT 4959
Db 1048 LysAsnLeuValLeuGlnAlaSer ----- 1055
QY 4960 AACTACAAATACCTTGAAGAATTGACGAGATGTGCTTTCTCTCTCTGCACTACGT 5019
Db 1055 ----- 1055
QY 5020 ATATGTTTCATCATGATGATGCAATTAATACAGATCCAGAAATTAAGATGAATGG 5079
Db 1055 ----- 1055
QY 5080 TTGCGAAAAAGATTGATGGAAGTGGCGGCAAGGAGGATCAATTTGGCTCAATCTAC 5139
Db 1056 -----GlySerThrGlu ----- 1059
QY 5140 AAGAAAAAAGAAAAAGAAAAATTAACATCGAGATCGCAACAATATCTTATGAG 5199
Db 1060 -----GlnGluLysGluLysMetLys ----- 1066
QY 5200 GTCCCGCTGTAGTGTATGAATATAGTTTATGATTAAAGATATATTCATAGT 5259
Db 1066 ----- 1066
QY 5260 ATTGATTAATTGGAGATGAAAAACAAAGACCGAGAAAAATTTGAAGAAATTTTAAAC 5319
Db 1067 -----GlnIleGlnAlaLysIleLysLysIleLeuAsn 1077
QY 5320 AAAAATGACATCAGTTGGCAAGAGATAGTACTACAGGAATCCCGAGTACT 5379
Db 1078 ---GlyAlaThrSer -----GlyValProProValThrLysAsnSerValLysThr 1093
QY 5380 GCGCGAAATTTTCTGCAAGCAAAATTAAGCAATGTGTGCAACGCAATGATGCGGG 5439
Db 1094 ProGlnGlnThrTrpGluAsnIleAlaLysAspIleTrpAsnAlaMetValCysAla 1113
QY 5440 ---TACAACGTGTAGGAGTATGATGAATAATGTGCAAAATGTGCAAGATGATGA 5493
Db 1114 LeuThrTyrLysGlu -----AsnAspAlaArg ----- 1122
QY 5494 GATCTAAAAAATGTGTCTGTACCTTACATGATGATATTCCTATGGGAAAAATCGC 5553
Db 1122 ----- 1122
QY 5554 GATGAAGTACTCGATATCAATTCTTCGATGGTTGCGCAATGGGTGAAGATTTTTCG 5613
Db 1123 ---GlyThrSer ----- 1125
QY 5614 AAACATTAAGAAAAAGAAATGAGAAATTTGTTAGGGCGCTGTAATGATTACTTGTGT 5673
Db 1125 ----- 1125
QY 5674 GATTAATGAAGATAAAGAAAGAAATGTACAGATGCGTGTACACAATTAATAAATTTAT 5733
Db 1125 ----- 1125
QY 5734 AGTAGAGTAACACACAGTATGAATAAACAATATAATATGTGAGAAATAAGACAAA 5793
Db 1126 ---AlaLysIleGlnGlnAsnLysAsp ----- 1133
QY 5794 ATATATTCCAGATCTCTGTGGCAAAAGATGACAGAGACCGCTGCCAATATTTTAGACAAA 5853
Db 1133 ----- 1133
QY 5854 CAATTAATAAATAATTTGTGAATAAATAAGTGAAGATTGTGAATATAGTGTATGAAGAT 5913
Db 1134 ---LeuLysLys ----- 1136

QY 5914 GTGTCCACACAGGATTAACTGATGTAAATAGTCAAAATATAGCCCGCATTAAGACAT 5973
Db 1137 ---AlaLeuTrpAspGluAlaAsnLysAsnThrPro ----- 1147
QY 5974 GAACCAAAAGAAAGTGAAGAAAGTGTAAATGTTCAGTGCACAGAGTCCACACAGTGA 6033
Db 1147 ----- 1147
QY 6034 CGAAGGGAACACCGTCCACACCGGTATCATGATATCAAAAGCAGACGCATCGAAAAA 6093
Db 1147 ----- 1147
QY 6094 GAAGCGAAAAACGCCCGCTACAAAAACCGCAAAAAAGTGGAAATTAACACAGAA 6153
Db 1147 ----- 1147
QY 6154 ATGGAGCACAAACAGAACCCGACGACGACACAAACAGAAACAGAACATCAACA 6213
Db 1147 ----- 1147
QY 6214 GCAACAAACACAGATCTGACGTGGGCAATGTGTAAAGCCATTTCTTGAATTAACCA 6273
Db 1147 ----- 1147
QY 6274 GATAGCAGGGGTGAATGAGAGGTGTAAATCCAAAAACGTATGACAAATATCCTAATGG 6333
Db 1147 ----- 1147
QY 6334 GGTGTATTTAGTAAAGTAAAGAAATGAAATGGCATATGATGCTCTAGAGAGA 6393
Db 1147 ----- 1147
QY 6394 AAAAATATGTAATTAATATATACATTTTAATTAATGAACTGAAATTAACCGTGAC 6453
Db 1148 ---IleGluLysTyrGlnTyrThrAsnValLysLeuGlu ----- 1159
QY 6454 AATGATTAATAAGAGCGCTTATTAATGTACAGCAATAGAACTCAATTTTGTGGTGA 6513
Db 1159 ----- 1159
QY 6514 AAATATTAATTAATAATCTCGACAGCAAAATGAATTCGAAATGGAACAATTCAGAT 6573
Db 1159 ----- 1159
QY 6574 GAATTTAAAGAAATATGATTAATACATATGATGATTAAGATATGTTTTCGAACT 6633
Db 1159 ----- 1159
QY 6634 GATATTTCTAATGATTAATAAATAATTAATTAATGTAACAATAAGTGTAAACCAATTCAT 6693
Db 1160 -----Asp ----- 1160
QY 6694 GAAATTAATTAAGAAAAACGATAAAAAAGATGAAGAAATTAAGTAAATTTTGG 6753
Db 1161 GluSerGlyAlaLysSerAsnAsp ----- 1168
QY 6754 GAGAAAAATTAATAATTTATTTGGAGAGATGATATATGATTAATCACTATCATCTACA 6813
Db 1168 ----- 1168
QY 6814 GACGAAAAAGAAAAAAGAAAAATTAAGATTAATTAACGATACATGATGACCAAACTG 6873
Db 1169 -----PheIleGln 1171
QY 6874 ACAGCTCCCTTGAAAGGTTGTAAAGAGCCCAATTTTGAATGCTTACACAGATGG 6933
Db 1172 ProThrLeuLysAsnProValGlnIleProThrPhePheArgLysGluTrp 1191
QY 6934 GCAGAAATTTGTATTAAGAGCAAGCAAGTGTAAATTAATGAGGGGGGGGTG --- 6990
Db 1192 GlyAsnSerPheCysPheGlnAlaGlnAlaLysArgLeuAlaGlnIleLysHisGluCysMet 1211


```

QY 2059 AAAGAACTTTCGAAATATATAAAGAAAAAATGAGTTTCCATTGGAATAT 2118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 577 GYASLHLEUENGLNLSNILEYASPVAlHISGLYASPhrAspRIleYSHISile 596
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2119 AGGACATTTAGAG-----AT 2136
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 597 LYSLSLEUENUSPGLUGLUSPAlAVALAlAVALleUGLYLYSAspAsnThr 616
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2137 GCATAGAACTCTGTAGACTTAAAGAACTCCAGAAATGTAAGCAATAT 2196
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 617 ThrlleAspLYleUENGLNHSGLUGLAlleUGLYSGLNLYS--- 635
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2197 ACAACGACACATGTGA-----ACATCCATAATGCAACACA 2235
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 636 ---GlnGluGluGlySGLYLSYAlAGLInGluGluSerArgLYASerAlaGLuThr 654
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2236 AACCCGCTGTAAACCTCGTGAAGCAGCACCCTAAATATAAAGAAATGCA 2295
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 655 ARGGLUSPGLUArgTYrGln-----GlnProAlaAspSerAlaGLYAluValGlu 671
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 2296 CAATACTTTAAAGAGTGCATACGAGAA 2325
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 672 GLUGLUGLUSPASPASPAspTYrAspGlu 681
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
US-08-487-826B-10
; Sequence 10, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-10

```

```

Alignment Scores:
Pred. No.: 2,36e-70 Length: 700
Score: 1032.50 Matches: 247
Percent Similarity: 50.28 Conserved: 110
Best Local Similarity: 34.79 Mismatches: 250
Query Match: 5.318 Indels: 103
Gaps: 22

US-10-087-013-1 (1-10628) x US-08-487-826B-10 (1-700)
QY 406 AATGATGCTTACCTCTGCGCCACCAAGACGACATGTGTGTAATAAATGGA 465
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 AsnLysValGlyAlaCysAlaProTYrArgArgLeuHISLeuYAspTYrAsnLeuGlu 24
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 466 GCCTTAATGATTAATTAATCCCAAAATATCATGATTATTGGAAATGACTAGTACA 525
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 25 SerIleAsp-----ThrThrSerThrHisLysLeuLeuGluValCysMetAla 42
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 526 GCAAAATACGAGAGTGAAATGTAATTAATCAT-----CCACATAAGGAACT-- 576
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 43 AlAlsTYrGluGluGlyAsnSerIle---AsnThrHisTYrThrGlnHISGlnArgTYrAsn 61
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 577 -----TCAGACGCTTGACTGCTTGCCAGAAAGTTTGACAGATTATAGTGAT 624
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 62 GluAspSerAlaSerGlnLeuCYSThrValLeuAlaArgSerPheAlaAspRIleGLYAsp 81
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 625 ATGTGAGAGCAATAGATATGTT-----AAACCAATGTCCATGAC 666
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 82 lIleValArgGlyLysAspLeuTYrLeuGlyTYrAspAsnLYSGLuGlnAluGlyLys 101
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 667 AAAGTACAAACGGGCTCCGAGAGTTTTCAGAAATATACATGATGAAATGAGAA 726
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 102 LysLeuGluGlnLYSLeuLYSAspRIlePheLYSLeuIleHis-----LysAspVal 118
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 727 GTAAAAATGAT-----TACAACTCATGGA---TCGCAAAATTATAT 768
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 119 MetLysThrAsnGlyAlaGlnGluArgTYrIleAspAspAlaLYSGLYAspPhePhe 138
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 769 AAATTAAGAGAGCATGCGTAATGTAATGAAATTAATGGAAGCTATTAACATGT 828
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 139 GlnLeuArgGluAspTrpTrpHisAsnArgGluThrValTrpLysAlaLeuIleCys 158
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 829 GATGCATCATTAATCTGATATTATTCGAATACAGAAAGTATACCATTAATTTCA 888
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 159 HisAlaProLYSAlaAsnTYrPheIleThrAlaCysAsnValGlyLysGlyThr 178
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 889 AATCTTAATGCGGCCCTAAACAAGCAAGCTCTTACCAATTAGATTGTCCCTCAA 948
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 AsnGlyGlnCysHisCysIleGlyGlyAspValProThrTYrPheAspTYrValProGln 198
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 949 TATTACGTTGGTTCGAGAAATGGGAGAGAGTTTGGCGAAATAATATTAAATG 1008
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 199 TyrLeuArgTrpPheGlnGluTrpAlaGluAspPheCysArgLYSLeuLYSLeu 218
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1009 AAAAAGTCAAGACCTCTGTAATGACAAGAACGTTATATGCTATCATTAATGA 1068
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 219 GluAsnLeuGlnLYSGLNAspTYrGluGlnAsnLeuTYrCysSerGlyAsnGly 238
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1069 CATGATGTACGACAACTATTGAAAAAGCATTTTGCATTGCTAATTAAGTACT 1128
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 TYrAspCysThrLYSThrIleTYrLysLYSLeuValIleGlyGlnHISCYSThr 258
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1129 GACTGTTGCAATAATGCAAGATTGGAAGTTTGAGGATTAACAAGAAAGCTTT 1188
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 259 AsnCYSerValTrpCysArgMetTYrGluThrTrpIleAspAsnGlnLYSGLuPhe 278
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1189 AAAAACAAGAAATAATGAAAAAGAAATA-----1221
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 279 LeuLYSGLNLYSArgLYSTYrGluThrGluIleSerGlyGlySerGlyLysSerPro 298
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 1222 -----CATCATATTATCGAAGCATATACAAATTGTCAT 1257
      |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 299 LysArgThrLysArgAlaIalArgSerSerSerSerAspAspN----- 314
QY 1258 AATATTAAAGTGAATATTATTAACAAATTTATGAAAACTTAAGGAAGCAATATGCA 1317
Db 315 -----GlyTyrGluSerLysPheTyrLysLysLysGluValGlyTyrGln 330
QY 1318 ACTAATGACACTTTTAAATTAATTAATGAAGAAAGTATGTAAAGCA----- 1368
Db 331 AspValAspLysPheLeuLysIleLeuAsnLysGluLysIleCysGlnLysGlnProGln 350
QY 1369 ---GGATTACAGAGAAAAGATATTACTTTACTTAACAGTGTGATGACAAAGGATA 1425
Db 351 ValGlyAsnGluLysAlaAspAsnValAspPheThrAsnGluLysTyrValLys---Thr 369
QY 1426 TTTTATGCTTGAATAATTCGCAAGTGTGCCGACGCGGAGTCAATGATGATGATA 1485
Db 370 PheSerArgThrGluIleCysGluProCysGluProTyrPheGlyLeuGluLysGlyGlyPro 389
QY 1486 AAATACACACCAATGAGTAAATGATCGTGAACGCTGAATTAATGAAGCTATAACCT 1545
Db 390 ProTyrPheValLysGlyAspLysTyrCysGlySerAlaLysThrLysThrTyrAspPro 409
QY 1546 CCATGGGGTGTGAAGCCCTAATATCATCTCTTATAGTGGTATGAACAAGTGGAT 1605
Db 410 Lys-----AsnIleThrAspIleProValLeuTyrProAspLysSerGlnGlnAsn 426
QY 1606 ATTACACAAAATTAAGAAATTTTGTACAGCTCACTAATTAACAAGATAAATATAT 1665
Db 427 IleLeuLysLysTyrLysAsnPheCysGluLysGlyValAlaPro---GlyGlyGlyGlnIle 445
QY 1666 CAAAATGCGAATGCTTATTAAGAT-----GAAATATAATTAAGTGT 1710
Db 446 LysLysTyrPheLysTyrTyrLysAspGluHisArgProSerSerLysAsnAsnAsnCys 465
QY 1711 AAACGTGACAAAATCTGAAATCAATATGATATCTAATGATATCATTCATTCATAT 1770
Db 466 ValGluGlyThrThrAspLysPheThrGlnGlyLysGlnThrValLysSerTyrAsnVal 485
QY 1771 TTTTGAATTAAGGTTACATTTATTAAGGATACATTAAGTGAATGACAAACTT 1830
Db 486 PhePheThrAspTyrValHisAspMetLeuHisAspSerValGluTyrPheThrGluLeu 505
QY 1831 AAACCTGTATTAATTAATACACACG-----CATTCGTATGATGAATGATAC 1878
Db 506 SerLysCysIleAsnAsnAsnThrAsnGlyAsnThrCysArgAsnAsnAsnLysCysLys 525
QY 1879 AGAATAGTGTATTTGTTGACGATGGTTAAACAAAAGAAAGAAATGATATGATA 1938
Db 526 ThrAspCysGlyCysPheGlnLysTyrValGluLysGlnGlnGluTyrPheMetAlaIle 545
QY 1939 AAGAACTGTGACAAAAGAAAGAAATTAACAGCAATGCTATTATAGTAAATTAATAT 1998
Db 546 LysAspHisPheGlyLysGlnThrAspIleValGlnGln-----LysGly 560
QY 1999 CTTTGTGAAGTTATTTTAAAGTTATGATTAACCTTGACAAAGATGACAAATGCG 2058
Db 561 IleIleValPheSerProTyrGlyValLeuAspLeuVal-----LeuLysGly 576
QY 2059 AAGAAGCTATGAGAAATATTAAGAAAAAAGAAAGTGTTCATTCATTTGGAATAATAT 2118
Db 577 GlyAsnLeuLeuGlnAsnIleLysAspValHisGlyAspThrAspAspIleLysHisIle 596
QY 2119 AGGGACTATTAGAG-----AAT 2136
Db 597 LysLysLeuLeuAspGlnGluAspAlaValAlaValIleLeuGlyGlyLysAspAsnThr 616
QY 2137 GCAATAGAACCTTTGTGATCACTTAAAGAAAGTGCACAGATATGTAAGACAATATAT 2196
Db 617 ThrIleAspLysIleLeuGlnIleGluLysGlnGlnIleGluGlnLysGlyGlnLys--- 635
QY 2197 ACAAGCAAGAGATGGA-----AATGCCAATATGCAACA 2235
Db 636 ---GlnGlnGluCysGlnLysLysAlaGlnGlnLysSerArgLysArgSerAlaGluThr 654

```

```

QY 2236 AACCCGTGCTTAACCTCGTGGAGCGACGCAACCCACTAATAAATTAAGAAATGACA 2295
Db 655 ArgGluAspGluArgThrGln-----GlnProAlaAspSerIleGlyGluValGlu 671
QY 2296 CAATACTTTAAAGAGATGATACGAGAA 2325
Db 672 GlnGlnGluAspAspAspAspTyrAspGln 681

RESULT 8
US-09-210-288-10
; Sequence 10, Application us/09210288
; Patent No. 6392026
GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Childs, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Kin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEO ID NO: 10:
SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-09-210-288-10

Alignment Scores:
Pred. No.: 2,36e-70 Length: 700
Score: 1032.50 Matches: 247
Percent Similarity: 50.28% Conservative: 110
Best Local Similarity: 34.79% Mismatches: 250
Query Match: 5.31% Indels: 103
DB: 4 Gaps: 22

US-10-087-013-1 (1-10628) x US-09-210-288-10 (1-700)
QY 406 AATGATGCTATACCTGTGGCCACCTAGAGAGCATATGTGATTAATAAAGTGAAG 465
Db 5 AsnLysValGlyAlaCysAlaLysProTyrArgArgLeuHisLeuLysAspLysLeuGlu 24
QY 466 GCCTTAATGATTAATTAATACCAAAATATTCATGATTTATGGGAAATGCTACTTACA 525

```



```

APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhan
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE: Plasmodium falciparum
US-08-568-459A-8

Alignment Scores:
Pred. No.: 3,63e-41
Score: 649.50
Percent Similarity: 39.48%
Best Local Similarity: 26.18%
Query Match: 3.34%
                2
                Gaps: 48

-10-087-013-1 (1-10628) x US-08-568-459A-8 (1-921)
QY 2842 GATATATTCGAGAGAGATCTCTGGAAGAAAGCGTGACATGGTAAAGTCGACAGCA 2901
DB 80 Asptylrlllethralaseraspile--GluasnlglylAsnserlle----- 94
QY 2902 CATTTGGAACGTTTGTGTAATATACATAAGTCACTCAAGCAAGCAAGATGATAA 2961
DB 95 -----GlyasnilleaspMetValIserAspIysAspAlasnlgly 108
QY 2962 TATTAAT-----GATGATGCCGCCCAATATTTAAATTTAGAGCAAAATTGG 3006
DB 109 PheasnnglyLeuaspAlacysglySerAlaasnIlePheIysglyIleargLysgluIn 128
QY 3007 TGGGAAGCTAATAGACCAAGTATGGAGAACCATGAATGTGATATAAATATTGANG 3066
DB 129 TrpIysCys--AlalysVal-----CysglyLeuaspValCysgly 141
QY 3067 GATATGCGGACACCAATCAACACAAAGTATTTGCGGATATAGTATACATACACA 3126
DB 142 LeuIysasnnglyAsn-----GlyserIleaspIysAspGln 153
QY 3127 TTGATGATATATCCACAAAAATTAGATGATGACCAAGTGGCAGATGCTACTGC 3186

```

```

DB 154 LysgluIlellelleIleargAlaleu-----LeuIysargTrpValglutyrPheleu 171
QY 3187 AAGTGCAGAAAAGAGTATGTAAGTGAAGAGAAAGTGAAGAGTGAAGTAAAG 3246
DB 172 -----GluaspIyrAsnIysIleasnAlalysIleSerHisCysThrIysLys 187
QY 3247 GATATGTCGACAGCGCTGTCGAAAAGAGTGTACAGAGTGTACAGAGTGTACAGAGT 3306
DB 188 AspasnIleSerThrCysThrAsnAspCysProasnLysCysThr--CysValglu--- 205
QY 3307 TGTATGATATATATATATATATATATATATATATATATATATATATATATATATAT 3366
DB 206 -----GluTrpIleasnnglyLysargTrpGluTrpIysasnIleLysLys 220
QY 3367 AATATCAAGAAATTCATGAA----- 3387
DB 221 HistIysThrGlnasnnglyLysasnnglyLysAspAsnMetLysSerLeuValThrAspIle 240
QY 3388 -----CAAGCAAAATGCTGTCTAGTAAAT-----AGTGTATAT 3420
DB 241 LeuGlyAlaleuGlnProGlnIserAspValasnLysAlaIleLysProCysSerGlyLeu 260
QY 3421 GAAGCT-----TTCAGTACGCCCAAAATCAT 3447
DB 261 ThrAlaPhegluSerPheCysglyLeuasnnglyAlaAspAsnSerGluLysgluIn 280
QY 3448 ATAGAC-----AGCAATGTATATATATATATATATATATATATATATATATATAT 3483
DB 281 GluaspIyrAspLeuValLeuCysMetLeuLysAsnLeuGluLysGlnIleGlnIleLys 300
QY 3484 TACCAACAAATATGTCGCAAAAGTAAATGATGATGATGATGATGATGATGATGATGAT 3543
DB 301 LysLysLysHisglyGluThrSerValgluasnngly-----GlyLysSerCysThrPro 318
QY 3544 GGTACTAACCCAGTATGAAATGTTGAGACATATCTCATGATACAGAAATTTGAT 3603
DB 319 LeuAspAsnThrThrLeuGluGlu-----Glu 327
QY 3604 GATGTCAGTCAACAAATAGAC-----TTTGTGATGAAAAAGT----- 3642
DB 328 ProIleGluGluGluAsnGlnValGlnAlaProAsnIleCysProLysGlnThrValGlu 347
QY 3643 GATGCTAAGATTAACGAAAA-----TATGCTTAAAGATTA 3681
DB 348 AspIysLysLysgluGluGluGluThrCysThrProAlaSerProValProGluLys 367
QY 3682 CCACAGACCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3741
DB 368 ProValProHis-----ValAlaArgTrpArgThrPheThrProProGluValPheLys 385
QY 3742 ATAAAAAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3801
DB 386 Ile-----TrpArgGlyArgArgAsnLysThrThrCysGluIleValAlaGluMet 402
QY 3802 CTTAAGAAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3861
DB 403 LeuIysAspIysasnnglyArgTrpThrValGlyCysGlyTrpArgLysgluThr----- 420
QY 3862 GATATCCGATTCGCAATCGCAATATATATATATATATATATATATATATATATATATAT 3918
DB 421 ---TysSerIleuTrpThrCysAspGluSerLysIleLysMetGlyGlnHisGlyAlaCys 439
QY 3919 ATGCCCCCTTGAAGACAAAGTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3978
DB 440 IleProProArgArgGlnLysLeuCysLeuHisTyrLeu-----Glu 453
QY 3979 AATTTCAATCAACAAATTAAT-----TTAAAGAGGCTTTCATCAAAATTCGACAGCAGAG 4035
DB 454 LysIleMetThrAsnThrAsnGluLeuLysTyrAlaPheIleLysCysAlaIleAlaGlu 473
QY 4036 ACATTCCTCTCATGTAT---TATTATAAAGTAAAGATGCGGAAGAAATGAAGTCAAT 4092
DB 4036 ACATTCCTCTCATGTAT---TATTATAAAGTAAAGATGCGGAAGAAATGAAGTCAAT 4092

```


Db 474 ThrPheLeuLeuTrrpGlnAsnTyrIlySAspIlySAsnGlyAsnAlaGluAspLeuAsp 493
 QY 4093 AAGAATTAAGAGAGCAAAATTCCTCCGATTTTGGATTCATGCTTACACATTT 4152
 Db 494 GluIlySLeuGlyGlyIleIleProGluAspPheIlySArgGlnMetHeterThrPhe 513
 QY 4153 GCGATTAATAGATTTTATTTTGGACAGATATA-----TCAAAA 4194
 Db 514 AlaAspTyrArAspIleCysLeuGlyThrAspIleSerSerIlySAspThrSerIlyS 533
 QY 4195 GGTCAATGGAGGAGAGTAAGTAAGCAAAATGATTCCTTCTTCAAAAATGGTGAC 4254
 Db 534 GlyValGly-----IlyValIlySAsnIleAspAspValPhe----- 546
 QY 4255 CAAAAATCTCTAATGA---AAACACGCCAGAAATGGTGACAGACATAGTCATGAG 4311
 Db 547 TyrIlySLeuSerAsnSerIleArgTyrArgIlySertPrIleGlnAsnGlyProVal 566
 QY 4312 ATATGGGAAGCTATGCTATGTCACATA-----GTAAAAAT 4347
 Db 567 IleTrrpGluGlyMetLeuCysAlaLeuSerTyrAspThrSerLeuAsnAsnValAsnPro 586
 QY 4348 GGGGCAAAAAAAGATGATTTTACCCGAAACTACGTTACACAGCTCAATTT---AGT 4404
 Db 587 GluThrHisIlySLeuLeuThrGluGlyAsnAsnAsnPheGluIlyValIlePheGlySer 606
 QY 4405 GACAAAAGCACACTTTGGAGGAATTTGCCAAACGACCCAGTTTTCAGATGGCTACAC 4464
 Db 607 AspSerSerThrLeuSerIlySPheserGluArgProGluPheLeuArgTrIleuThr 626
 QY 4465 GAATGGTACGACACTATTCCTATACACACAAAAATTTGAGAGATGTCACGAAAAA 4524
 Db 627 GluTrrpGlyGluAsnGlySylGluGlyIlySValLeuLeuAlaIlyS 646
 QY 4525 TGTAAATCAATGACCAATGTAAGTGTATACAGATGTAATAAGAA-----TGC 4575
 Db 647 CysIlySAspCysAsp---ValAspGlyAspGlyIlySAsnGlySylValAlaCys 665
 QY 4576 GAGGACTACGTTAA-----TATATGAAA 4599
 Db 666 LysAspGluCysIleGlySerThrIlySertPrIleGlyIleTrrpLeuAspAsnTyrIlyS 685
 QY 4600 AAAAAAAGATGCGATTCGCACAGATTA-----TATTACAGATGACAGCGCAAA 4653
 Db 686 GlnIlySGLyArgTyrThrGluValIlySLeuIleProLeuTyrIlySGLyAspVal 705
 QY 4654 AAAAGATTGAT-----AGACACACATTTGGTGAATGTTACAGACTATACGAGAC 4707
 Db 706 LysAsnSerAspAspAlaArgAspTyrIleuIlySThrGlnLeuGlnAsnMetIlySVal 725
 QY 4708 AATGCAACAGATTACTTGAAC-----AGCAAAATTTACTGCTACTGTG 4749
 Db 726 AsnGlyThrThrAspGluAsnGlySylGlyIlySAsnMetHleuArgHeterThrAsn 745
 QY 4750 GGTGATTAACCTGGAGAGTGCCTGTGTGTACAAAGAAATATACAAATTTGAGAAAAACG 4809
 Db 746 SerAspMetProGluSer-----LeuAspGluIlyS 755
 QY 4810 GCTTACTATGATGCGCAAAACATTTGGTGACAAAAATTTATGAAATGACGACAAA 4869
 Db 756 ProGluIlySValIlySAspIlyS---CysAsnGlyS---ValProAsnGluCysAsn 771
 QY 4870 TATACTACATTTGAGT----- 4887
 Db 772 AlaLeuSerValSerIlySerglyPheProAspGlyGlnAlaPheGlyGlyValLeu 791
 QY 4888 AAGATTAAGTCAAGAGATTAAGTAAGAGGCAACAGGTCTTAAAGTGGCAAAAC 4947
 Db 792 GluGlyThrCysIlySGLyLeu-----GlyGluProIlyS---LysIlyS 804
 QY 4948 AAGGTCTCAATACTACATACTTGAAGAATGACTGAG----- 4990
 Db 805 IleGluProGluGlnTyrAspProThrAsnAspIle-LeuIlySertThrIleProValTh 824

QY 4990 ----- 4990
 Db 824 rIleValLeuAlaLeuGlySerIleAlaPheLeuPheMetIlySValIleTyrIleTyrVa 844
 QY 4990 ----- 4990
 Db 844 lrrpTyrIleTyrMetLeuCysValGlyAlaLeuAspThrTyrIleCysGlyCysIleCy 864
 QY 4991 -ATGCTCTTTCCTCTCTGTCGAC-----TAGCAATATGTTTCATGCA 5034
 Db 864 sIleCysIlePheIleCysValSerValTyrValCysValTyrValTyrValPheLeuTyr 884
 QY 5035 T-----TGATGGCAATTATACAGATCCAGATTAAGATGAAAAAT 5076
 Db 884 rMetCysValPheTyrIleTyrPheIleTyrIleTyrValPheIleLeuIlyMetIlySly 904
 QY 5077 GGGTTCGCAAAAAGATTGATGGAAGTGGCGCAACGGAAGGTTACATTTGGCTCAATAC 5136
 Db 904 s-----MetIlySly 907
 QY 5137 TCAAGAAAAAAGAAAAAGAAAAATMAAAAGCTGGAG 5179
 Db 907 sMetIlySlyMetIlySlyMetIlySlyArgIlySlyArgIly 921
 RESULT 10
 US-08-487-826B-8
 ; Sequence 8, Application US/08487826B
 ; Patent No. 5993827
 GENERAL INFORMATION:
 APPLICANT: Sim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CP1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 921 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 US-08-487-826B-8

Db 188 AspAsnGluSerThrCysThrAsnAspCysProAsnLysCysThr---CysValGlu--- 205
 QY 3307 TGTATGAAATATATGATTAATATGATTTATGAAAGACAAATGAAATATATATATGAT 3366
 Db 206 -----GluTrpIleAsnGlnLysArgThrGluTrpLysAsnIleLysLys 220
 QY 3367 AAATACAAAGAAATTACATGAA----- 3387
 Db 221 HisTrpLysThrGlnAsnGlnAsnGlnLysAspAsnAsnMetLysSerIleuValThrAspIle 240
 QY 3388 -----CAAGACCAAAATGCTGTAGTAAT-----AGTGTAT 3420
 Db 241 LeuGlyAlaLeuGlnProGlnInsSerAspValAsnLysAlaIleLysProCysSerGlyLeu 260
 QY 3421 GAAGCT-----TCCAGTACTGCCCAAAATCAT 3447
 Db 261 ThrIleAspGluSerPheCysGlyLeuAsnGlyAlaAspAsnSerGluLysLysGly 280
 QY 3448 ATGAGC-----ACGAATGTTAATGAAATTTTGTCCGAATTA 3483
 Db 281 GluAspTrpLysLeuValLeuCysMetLeuLysAsnLeuGlnLysGlnIleGlnGlnLys 300
 QY 3484 TACCAACAAATGTTGGCAAAAGTAATAAAGTGTACTAGTACGAAAGTGTCTATT 3543
 Db 301 LysLysLysHisGlyGluThrSerValGluAsnGly-----GlyLysSerCysThrPro 318
 QY 3544 GGTACTAACACCAAGTAAATGTTGACACATATCTCATGATACAGAAATTTGAT 3603
 Db 319 LeuAspAsnThrThrLeuGlnGlu-----Glu 327
 QY 3604 GATTGTCACTGCACAAATGAG-----TTTGTGATGAAAAAGT--- 3642
 Db 328 ProIleGlnGlnGlnAsnGlnValGlnAlaProAsnIleCysProLysGlnThrValGlu 347
 QY 3643 GATGTTAGATTAACGAAAA-----TATGCTTTAGAGATAA 3681
 Db 348 AspLysLysLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 367
 QY 3682 CCACAGACACATGATGCTGTGCTGTGTAAGAGTGAACCGCAACCGCACAGGATCAG 3741
 Db 368 ProValProHis-----ValAlaArgTrpArgThrPheThrProGluValIlePheLys 385
 QY 3742 ATAAACAAAGAAAAAGCGGAAAGAAAGACATACGAAATGTAACACATGATGATTA 3801
 Db 386 Ile-----TrpArgGlyArgGlyAsnLysThrCysGlnLysGlnLysGlnLysGlnLys 402
 QY 3802 CTTAAAGAAACGATGGAAGAAAGAAAGATGATGATGATGATGATGATGATGATGAT 3861
 Db 403 LeuLysAspLysAsnGlyArgThrThrValGlyLysCysGlyArgLysGlnLysGlnLys 420
 QY 3862 GGATATCCCATGCGCATGCGGAAATATTAATTA---GTGGAAGACCTCGTGTGT 3918
 Db 421 ---TyrSerIleTrpThrCysAspGluSerLysIleLysMetGlnHisGlyAlaCys 439
 QY 3919 ATGCCCTTGAAGACAAAGTATGCGACATTTCTTGCAAAATGATTAAGTAATAA 3978
 Db 440 IleProProArgGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 453
 QY 3979 AAATTTCATTCACAACTTAAT---TTAAAGAGCTTTCATCATCAATTCGACAGCAGAA 4035
 Db 454 LysIleMetThrAsnThrAsnGlnLeuLysGlyAlaPheIleLysCysAlaIleAlaGln 473
 QY 4036 ACATCTCTCATGCTAT---TATTATTAAGATGATGATGATGATGATGATGATGAT 4092
 Db 474 ThrPheLeuLeuTrpGlnAsnGlyLysLysLysLysLysLysLysLysLysLysLysLys 493
 QY 4093 AAAGATTAAGAAAGCAAAATTCCTCCGACATTTTGAATGATGATGATGATGATGAT 4152
 Db 494 GluLysLeuLysGlyLysIleIleProLysAspPheLysArgGlnMetPheTyThrPhe 513
 QY 4153 GGAGATTATAGAGATTTTATTGGAACAGATATA---TCAAAA 4194
 Db 514 AlaAspTrpArgAspIleCysLeuGlyThrAspIleSerSerLysLysAspThrSerLys 533

QY 4195 GGTATGTTGAGGAGAACTAAACTAAAGACCAATAGATTCTCTTTCAAAATGGTGAC 4254
 Db 534 GlyValGly-----LysValLysCysAsnIleAspAspAlaPhe----- 546
 QY 4255 CAAAATCTCTTAATGGA---AAAACACCGCAAGATGGTGCAGACAAATGATGAG 4311
 Db 547 TyrLysIleSerAsnSerIleArgTrpArgLysSerTrpTrpGlnThrAsnIleProVal 566
 QY 4312 ATATGGAACTATGCTATGTCACATA-----GTAAAAAT 4347
 Db 567 IleTrpGlnLysMetLeuCysAlaIleLeuSerTyrrAspThrSerLeuAsnValAsnPro 586
 QY 4348 GGGGCAAAAAAGATGATTTTACCGAAACTACGTTACACACGTCGTAATTT---AGT 4404
 Db 587 GluThrHisLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 4404
 QY 4405 GACAAAAGCAACCACTTGGAGGAATTTGGCAACGCCAGTTTTCATGATGCTAAC 4464
 Db 607 AspSerSerThrThrLeuSerLysPheSerGluArgProGlnPheLeuAlaGlyLeuThr 626
 QY 4465 GAATGTCACGACGACTATGCTATACAGACAAAATTTGAAGATGTGCAGGAAAA 4524
 Db 627 GluTrpGlnGlnAsnPheCysLysGlnGlnLysLysGluTyrrLysValIleLeuAlaLys 646
 QY 4525 TGTAACTCAATGACCAATTTGAAGTGTGATACAGATGATATAGAAA-----TGC 4575
 Db 647 CysLysAspCysAsp---ValAspLysAspLysCysAsnGlyLysCysValAlaCys 665
 QY 4576 GAGGACTACGTTAA-----TATATGAAA 4599
 Db 666 LysAspGlnCysLysGlnTyrrHisSerTrpIleGlyIleTrpIleAspAsnTyrrLysLys 685
 QY 4600 AAAAAAAGAGATTCACAAAGATAA-----TATTAACAAAGATGACCGCAAA 4653
 Db 686 GlnLysGlyArgTyrrGlnValLysLysIleProLeuTyrrLysGlnLysAspVal 705
 QY 4654 AAAAGATTCGAT-----AGACAAACATTTGGTAAAGTGTACAGACATATCGAAG 4707
 Db 706 LysAsnSerAspAlaArgAspTyrrLeuLysThrGlnLeuGlnAsnMetLysCysVal 725
 QY 4708 AATGCAACAGATTAATGTAAC-----AGGAAATTTCTCTGATGCT 4749
 Db 726 AsnGlyThrAspGlnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 745
 QY 4750 GGTATTAAGCTGGAAGTGGCTGTGTCGTCACAAAGAAATATACAAATTTAGAAAAAG 4809
 Db 746 SerAspMetProGluSer-----LysAsnLys 755
 QY 4810 GCTTACTATGATCCGCAACAAATTTGGTGCACAAAATTTATGAAATGACGACAAA 4869
 Db 756 ProGlnLysValLysAspLys---CysAsnCys-----ValProAsnGlnCysAsn 771
 QY 4870 TATACATAACTTTCCGCT----- 4887
 Db 772 AlaLeuSerValSerLysSerGlyPheProAspGlyGlnAlaPheGlyLysValLeu 791
 QY 4888 AAAGATTAAGTGAAGAGTATTAAGAGGCAACACAGCTGCTATTAGTGGCAAAAC 4947
 Db 792 GluGlyThrCysLysGlyLeu-----GlyGluProLys---LysLys 804
 QY 4948 AAAGCTCTAATTAATCAATTAATGAAAGATTTGACTGAAG----- 4990
 Db 805 IleGluProProGlnTyrrAspProThrAsnAspIle---LeuLysSerThrIleProVal 824
 QY 4990 ----- 4990
 Db 824 IleValIleuAlaLeuGlySerIleAlaPheLeuPheMetLysValIleTyrrIleTyrrAs 844
 QY 4990 -----TCAAAA 4990
 Db 844 ITPTrIleTyrrMetLeuCysValGlyAlaLeuAspThrTyrrIleCysGlyCysIleCys 864

OY 4991 -ATGCGTTTTCCTCCTGCAGC-----TAGCTATTGTTCATGCA 5034
:::|||||::: ||||| |
Db 864 silecysilhephelecysvalserValtyrValcysvaltyrValtyrValphenleuty 884

OY 5035 T-----TGATGGCAATTATACAGATCCACAAGCTTAAGAATGA 5076
:::|||||::: ||||| |
Db 884 tmetcysValphenlyrphenlyrphenlyrphenlyrphenlyrphenlyrphenlyr 904

OY 5077 GGGTTGCCAAAAAGATTGATGGAAGTGGCGCAACGGAAGGTACAAATTTGGCTCAATAC 5136
:::
Db 904 S-----Metlysly 907

OY 5137 TACAAGAAAAGAAAAGAAAAGAAAATATAAACCTCGCATG 5179
||||| ||||| ::||| |||||:::|
Db 907 smetlyslysmetlyslysmetlyslysalglyslylsargille 921

RESULT 12
US-08-568-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306

```
QY 4332 TGCACATAGTAAATTTGGCCAAATAAGATGATTTTACCGAAACATGCTTACAAACA 4391
Db 301 -----
QY 4392 CGTCAATTTAGTACAAAAGCACCACTTTGGAGAAATTTGCCAAAGACCCGATTTT 4451
Db 306 ntleasn---AsnGlySerAsnIleProGlnGlnGluLeu-----GlnIleth 321
QY 4452 ACGATGCTTACCGAATGAGACGACGACATGCTTATACGACGCAAAAATTTGAAAGA 4511
Db 321 rGlnIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4511
QY 4512 TGTGACGAGAAATTTAGTACCAATGCAATTTGAAG---TGTGATACAGAAATTTAATA 4568
Db 341 uProGlySerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4568
QY 4569 GAATGCGAGGATGCTTAAATAT---ATGAAAAAATTTGATGCTTCCACACA 4625
Db 361 pProGlySerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4625
QY 4626 TAAATATTACAGAGTAAAGCGACAAATAAGATTGATGACAAACATGCTGCTAAT 4685
Db 381 rLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4685
QY 4686 GGTATACAGACTATCTGACGACGACGACGACGACGACGACGACGACGACGACGACG 4745
Db 393 -----
QY 4746 TTTGCTGATTAAGCTTGAAGTCCCTCTGCTGATACAAAGAAATTTACATTTGAAAA 4805
Db 404 nLysAsnAsp-----AlaLysValSerLeuLeuAsn 415
QY 4806 AAGAGCTTACTATGATCCGACAAACATTTGGTGC-----ACAAATTTATTGA 4856
Db 415 nAsnGlySerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4856
QY 4857 A-----AATGAC-----GACAAATTTACTAATTTCCAGTAA 4889
Db 435 sSerValLeuAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4889
QY 4890 AGATTAAGTGCAAA---GGATTAGTAAGAGAGCA---AACACAGTGCATTAATGCTCA 4943
Db 455 pAspSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 4943
QY 4944 AATCAAGTCTTAACTACAAATACTTGAAGAATTTGATGAGATGCTTTTCC 5003
Db 475 uGlySerAsnProGlyLeuLeuSer-----ThrLysAspValGlyValPr 490
QY 5004 TTTCTGCTGACACTATATATGTTTTCATGATGATGATGATGATGATGATGATGATGAT 5063
Db 490 oProLArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5063
QY 5064 TAAAGATGAAATGGGTTG---CGAAAAAGATTGATGAGAGTGGCGGACGAGAG 5117
Db 506 ---AspLysAsnLeuLeuMetIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5117
QY 5118 GTACAAATTTGGTCAATCTACAAAGAAAAAAGAAAAAAGAAAAAATTTAAAGCTGCGA 5177
Db 524 rArgIleLeuLysArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5177
QY 5178 TGGCCACAAATATTCTATGAGGTCCCGCTGTGATGCTGCTGATGCTGCTGATGCTGCT 5237
Db 539 -----ValCysLysIleLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 5237
QY 5238 TTTTAAGCATTTAATTTAGTATGATTAATTTGGAAGTAAAGAAAAAGACGAGCA 5297
Db 549 pIleArgAspIleIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5297
QY 5298 AATTTGAGAAAAATTTATACAAATAATGAAACATCTGCTGCGAAAGAGTGTATCTAC 5357
Db 566 gLysLeuValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5357
QY 5358 TACAGAAATTTCCCGTAGTACGCGCAAAATTTTCTGGAACGAAATTAAGAAATGTGT 5417
Db 586 -----AspLysLeuPheArgAspGlnIlePheLysValIleLysLysAspVal 601
QY 5418 GTGAACGCAATGATATCGGGTACAAACGCTGTGAGGATGATGGAATTTATGGAATAG 5477
Db 601 lTrpAsnValIleSerTrpValPheLys----- 610
QY 5478 TGCAGAAAGTATGAAGATCTTAAAAAATTTGTTCTGTACCTTCAGATGATATCC 5537
Db 611 -----AspLysTh 613
QY 5538 TATGGGAAAAATTCGATGAAGTACGCTGATCAGTTTCTTGATGCTTGGCGAATG 5597
Db 613 rValLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5597
QY 5598 GGGTGAAGATTTTTCGAAACATTAAGAAAGAAATTTGGAAGTGGAGGCGGTGAA 5657
Db 633 pGlyAspAspLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5657
QY 5658 TGATTAATCTGTGTATGATGAAGATTAAGAAAGAAATGTCAGATGCTGTACACA 5717
Db 653 sGlnLysProGlySerGlnAspAsp-----AsnGlySerLeuGlnGlnGlnGlnGlnGlnGln 5717
QY 5718 ATATAAAAATTTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5777
Db 668 rTyrLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5777
QY 5778 TGAC-----AATTAAGCAAAATTAATTTCCGACATCTGTGCAAAAGATGC 5825
Db 688 nGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 5825
QY 5826 AGAGAGCGCTGCGGATTAATTTAGCAAAACATTAATAATTTGTAATAATTAAGTGG 5885
Db 708 uValLysLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 5885
QY 5886 AGATTGTGAATATAG-----TGT----- 5904
Db 723 eGlnAspGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 5904
QY 5905 ---ATGAAGATGCTGTCACACAGCATTAATGATGATGATGATGATGATGATGATGATGAT 5960
Db 743 oGlnValLysAspValProIleSer-----IleIleArgAsnAsnGlnGlnGlnGlnGlnGlnGln 5960
QY 5961 ATCATTTAGACATGACCAAAAGATTT-----GAGGAAAGTATATGTCA 6008
Db 762 uAlaValProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 6008
QY 6009 AGTCCACGAGGTCACACAGCTGTACGAGGAAACACCGTCAACAGGCTATCACTGAT 6068
Db 782 uGlyProLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 6068
QY 6069 ATCAAAAGCAGGCAATCGAAAAAAGAGGAAACAGCGCCGCTTACAAACAGCCGAA 6128
Db 797 uSerLysIleSerValSer-----ProGlnAsnSerArgPro-- 809
QY 6129 AAAAGTGAATAATCTAACACAGAAATGCGAGCACAACAGAACCCGACGACGACACA 6188
Db 810 -----G1 810
QY 6189 ACAAAACGAAAAAGCAATCAACACACACACACACACACATGCTGAGTGGC---ACAAT 6245
Db 810 uTrpAspAlaLysAspTrpSerAsnLeuLeuLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 6245
QY 6246 GGTAAAGCCATCTTTCGATTAACCAATGACAGGCGGTGAATGAGGCTGTGATATCC 6305
Db 830 tProLysAlaValIleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 6305
QY 6306 AAAAGCTATGACAAATATCTAATGAGGCTGTATTTAGTATGATGATGATGATGATGATGAT 6365
Db 845 -----ValThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 6365
QY 6366 AATTGCAATATGATGCTCTCTAGAGAAAAAATTAATGATTAATTAATTAATTAATTT 6425
```

```

Db      852 eserGlyVal-----AsnSerLysProLeuSerAspAspValArgProAs 867
QY      6426 AAATTTAAGAACTGAAATTAAGCGGTGACATGATATAAAGAGCGTTTATTAATGTC 6485
Db      867 pLysLysGluLeuGlnAspGlnAsnSerAspGluSerGluGlnThrValVal----- 884
QY      6486 AGCAATATGAAATCTCAATTTTGTGGTTAAATATATATATGAAATATCCGACAGAAAA 6545
Db      885 -----AsnHisIleSerLysSerProSerIleAsnAs 895
QY      6546 TGAATTTGCAAAATGAAACATTCACAGATGAATTTAAAGAAATATGATATATACATATGG 6605
Db      895 nGlyAspAspSerGlySer-----GlySerAlaThrValSe 907
QY      6606 TGAATTAAGAATATGTTTGTGAACGTGATTTTCAATGATATAAATAATATTAACGTG 6665
Db      907 rGluSerSerSerSerAsnThrGlyLeuSerIleAspAspAspArgAsnGlyAspThrPh 927
QY      6666 AACAAATAGTGTACACACATCTCTCAATGAAATTAATAGCAAAACAGAGATAAAAAAA 6725
Db      927 eValArgThrGlnAspThrAlaAsnThrGlnAspValIleArgGlyGlnAsnAlaAspLys 947
QY      6726 AGATGAAGAATTAACGTAAATATTTTGGGAGAAATTAATAATTTATTTGGAGAGAAAT 6785
Db      947 sAspGluAspGluLysGlyAlaAspGluLysArg----- 958
QY      6786 GATATATGATTAATCATCATCTCACAGACGAAAC-----GAAAGAGAAA 6833
Db      959 -----HisSerThrSerGluSerLeuSerProGluLysSme 972
QY      6834 AATTAAGATTAATTAACAGTACATACATACATACGACAACTGACGCTCCCTGTAAGAGTT 6893
Db      972 tLeuThrAspAsnGluGlyGlyAsnSerLeuAsn-----HisGln 986
QY      6894 TGTAAAGAGCGCCCAATTTTGAATGGTTCACAGAAATGGCAGAAAGATTTTGTAAATA 6953
Db      986 uValLysGluHisThrSerAsnSerAspAsnValGlnGlnSerGlyGlyLeuValAsnSme 1006
QY      6954 GAGAGAGGAAACACTGTTTAA-----TTGAGGCGG 6986
Db      1006 tAsnValGluLysGluLeuLysAspThrLeuGlnAsnProSerSerSerLeuAspGluG 1026
QY      6987 CTGTAAAGATTAATGAGTATGATGATAT----- 7017
Db      1026 yLysAlaHisGluGluLeuSerGluProAsnLeuSerSerAspGlnAspMetSerAsnTh 1046
QY      7018 -----GAGGTAAGACACAGAAATGTGCAGAGCGCTGTATACATATCAAA 7064
Db      1046 rProGlyProLeuAspAsnThrSerGluGluThrThrGluArg-----IleSerAsnAs 1064
QY      7065 TTTTATTAAGAACTGAAATTAAGCAAAAGCAAAAGTTCACAAAGAT-- 7122
Db      1064 nGluTyLysValAsnGlnArgGlnAspGluArgThrLeuThrLysGluTyLysAspL 1084
QY      7123 -----AAAGATGCCAAAGATATAGAGATTAATCC 7151
Db      1084 eValLeuLysSerHisMetAsnArgLysSerAspArgLysLeuTyLysArgLys----- 1102
QY      7152 TTCTACTGAAGAAGACATAGAGAAGCAACATGCTCATGATATTTAAACATGAATT 7211
Db      1103 -----AsnSerAspLeuSerThrValAsnAspGluSerGluAspAlaGlnAlaLysSme 1120
QY      7212 AAAGAATTAATGTGCAATTAAGATGTTCTTGTATGCAAAACCTTTCACACATACC 7271
Db      1120 tLys-----GlyAsn-----AspThrSerGluMetSerHisAsnSerSerGlnHisI 1136
QY      7272 AAAAACAACAACAATCACAATCATCCGATGCTAATGATATGCCAGATCGCTGAT-- 7329
Db      1136 eGluSerAspGlnGlnLys-----AsnAspMetCysThrValGlyAspLe 1151
QY      7330 -----TATGTTCTGTAAAGATTTAAACAGTGTGAGTGTCCGTAATTTCAAAA-- 7377

```

```

Db      1151 uGlyThrThrHisValGlnAsnGluIleSerValProValThrGlyGluIleAspLys 1171
QY      7378 -----AAGGATCTATGATTCATACAAAAAATTAAGACCTAAATACC 7424
Db      1171 sLeuArgGluSerLysGluSerLysIleHis-----LysAlaGluGluLysArgLeuSe 1189
QY      7425 TATGAATTTGTATGAGAAACAGCATATTAATTTATCTAAAGAGCAAAATTAATATGGA 7484
Db      1189 rHisThrAspLysIleHisLys-----IleAsnProGluAspArgAsnSerAsnTh 1205
QY      7485 TATTACCTGAGAGCAAAATTTTACTATATGAGTCTACAAAGCAAAAGCAATTAATGGA 7544
Db      1205 rLeuHisLeuLysAsp-----IleArgAsnGluGlnAsnGluLys 1218
QY      7545 TAGTTGCACTAATTAATTAAT 7563
Db      1218 gHisLeuThrAsnGlnAsn 1224

RESULT 13
US-08-487-826B-4
; Sequence 4, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhuan
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1435 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-4

Alignment Scores:
Pred. No.: 1,09e-38 Length: 1435
Score: 618.00 Matches: 339
Percent Similarity: 36.80% Conservative: 223
Best Local Similarity: 22.20% Mismatches: 538
Query Match: 3.18% Indels: 428

```


Db 566 glyleuValGlyLysIleAsnThrAsnSerLysThrValHisAlaGlnLysLysAsn-- 585
 QY 5338 TACAGAAATCCGGTAGTACTGCCGGAATTTTCTGGAACGAATAATAGCAATGTCT 5417
 Db 586 -----AspLysLeuPheArgAspLysPheValIleLysLysAspVal 601
 QY 5418 GTGGAACCAATGATATGCGGCTACAAACGTTGGTGGATGATGAAATATGGAATATG 5477
 Db 601 LTPASNTValIleSerThrPheLys----- 610
 QY 5478 TGCAGAAAGTGAAGATCTAAAAAATGTGTTCTGTACTTCAGATGATGATTATCC 5537
 Db 611 -----AspLysThr 613
 QY 5538 TATGGGGAATAATCCGATGAAGTACTGCGTATGACTTTCTTCGATGGTTTCCGATG 5597
 Db 613 rValCysLysGlnAspAspIleGlnAsnIleProGlnPheArgGlyPheSerGlnThr 633
 QY 5598 GGGTGAAGTTTTTTCGAACATTAAGAAAGAAATGGAAATGGTAGGGCGCTGTA 5657
 Db 633 pGlyAspAspLysCysGlnAspLysThrLysMetIleGlnThrLysValGlnCysLys 653
 QY 5658 TGATTAATCTTGCTGATTAATGAAGATTAAGAAAGAAATGTACAGATGCGTGTACACA 5717
 Db 653 sGlnLysProCysGlnAspAsp-----AsnCysLysSerLysCysAlaAsn 668
 QY 5718 ATATAAAAAATTTATAGTACTGAGTAACACACAGTATGAAAAACAATCAAAATATG 5777
 Db 668 rTyrlLysGlnThrLysSerLysLysGlnLysLysGlnLysLysGlnLysGlnLys 688
 QY 5778 TGAG-----AATAAGACAAATATATCCGAGCATCCTGTGGCAAAAGATGC 5825
 Db 688 nGlnLysGlnLysGlnLysAsnThrLysMetLysSerLysLysSerLysLysProGln 708
 QY 5826 AGAGCAGCGTCGCAATATTTAGACAAACAAATTAATAAAATTTGTAATAAAAGTGG 5885
 Db 708 uValLysLeuLysLysLysSerGlnLys-----CysSerAsnLeuAsn 723
 QY 5886 AGATTGTGAATTAAG-----TGT-- 5904
 Db 723 eGlnLysGlnPheLysGlnLysLeuHisSerAspLysAsnLysCysThrMetCysPr 743
 QY 5905 -----ATGAAGATGTGTCACACACAGCATTAAGTATGATCAAAATATGCCCC 5960
 Db 743 oGlnLysLysAspValProIleSer--IleLeuArgAsnAsnGlnLysThrSerGln 762
 QY 5961 ATCATTAAGCATGAACCAAAAGAGTT-----GAAGAAAGTGAATTTGTC 6008
 Db 762 uAlaValIleProGlnLysAsnThrGlnLysLeuHisArgThrGlnLysProSerIleSerGln 782
 QY 6009 ACTGCCACGAGGTCCACACACGTTGACGAAGGAAACACCGTCACACAGGGTATCACTGAT 6068
 Db 782 uGlyProLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 797
 QY 6069 ATCAAAAGCAGGAGCATGGAAGAAAGAAAGAAAGAAAGAGCGCCCTACAAAACGCCGA 6128
 Db 797 uSerLysLysSerValSer-----ProGlnAsnSerArgPro-- 809
 QY 6129 AAAAGTGAATAATCTAACAAACGAATGCGACACAAACACGACCCGACGACGACACA 6188
 Db 810 -----Gln 810
 QY 6189 ACAAAACAGAAACAGACATCAACAGCAACAACAACAGATGTGACGTGGC--ACAAT 6245
 Db 810 uThrAspAlaLysAspThrSerAsnLeuLysLysLysLysLysLysLysLysLysLys 830
 QY 6246 GGTAAAGCCATTTCTTTCGATTAACAGATAGCAGGGGTGGAATGAGAGGTGTGTAATCC 6305
 Db 830 tProLysAlaValIleGlySerSerProAsn--AspAsnIleAsn-- 844
 QY 6306 AAAAAGCTATGACAAATATCTAAATGGGGTGTATTTAGCTAAAGTCAAAAGAAATGA 6365
 Db 845 -----ValThrGlnGlnGlnLysAspAsnIle 852

QY 6366 AATGGCATATGTATGCTCTCAGAGAAAAAATTGATTAATATATACAAATATT 6425
 Db 852 eSerGlyVal-----AsnSerLysProLeuSerAspAspValArgProAs 867
 QY 6426 AATATATGAACATGAAAAATACCGTCACATATATATAAGAGCGCTTTATTAATGTCC 6485
 Db 867 pLysLysGlnLeuGlnAspGlnAsnSerAspGlnSerGlnLysThrValVal----- 884
 QY 6486 ACCAATAGAACTCAATTTTGTGTTAAATATATATGAAATATCTGCACAGCAAAA 6545
 Db 885 -----AsnHisIleSerLysSerProSerIleAsnAs 895
 QY 6546 TGAATTCGAATAATGGAACAATTCAGATGAATTAAGAAATATGATTAATACATATG 6605
 Db 895 nGlnAspAspSerGlySer-----GlySerAlaThrValSer 907
 QY 6606 TGATTAATAAGATATGTTTTTGGAACTGATTTCTTAATGATTAATAAATATTAACGT 6665
 Db 907 rGlnSerSerSerSerAsnThrGlyLeuSerIleAspAspAspArgAsnGlyAspThrPh 927
 QY 6666 AACAAATAGTGTAAACACCATCTCAATGAATAATATAAGAAAGAAAGAGATAAAAAAA 6725
 Db 927 eValArgThrGlnAspThrAlaAsnThrGlnLysPheValIleArgLysGlnAsnAlaAspLys 947
 QY 6726 AGATGAAGATTAAGTAAATATTTTGGAGAAATAATTAATTTATTTGGAAAGCAAT 6785
 Db 947 sAspGlnAspGlnLysGlnLysAlaAspGlnLysLys----- 958
 QY 6786 GATATATGATTAATCACTATCACTCAGACAGCAAAAC-----GAAAAAGAAA 6833
 Db 959 -----HisSerThrSerGlnSerLysSerProGlnLysLys 972
 QY 6834 AATTAGATTAATTAACCACTACATGACATGACCAAACTGACGCTTCCTTGAAAGAT 6893
 Db 972 tLeuThrAspAsnGlnGlnLysGlyAsnSerLeuAsn-----HisGlnLys 986
 QY 6894 TGTAAAGGCCCCAATTTTGGATGTTGACAGATGTTGACAGATGGCAAGAAATTTGTATTA 6953
 Db 986 uValLysGlnLysIleThrSerAsnSerAspAsnValGlnGlnSerGlyLysIleValAlaAsn 1006
 QY 6954 GAGGAAGCAACGTTGTTAAA-----TTGAGGCGGG 6986
 Db 1006 tAsnValGlnLysGlnLysLysAspThrLeuGlnLysnProSerSerLysLysPheLys 1026
 QY 6987 CTGTAAGCAATTAAGTGTATGATGTTAGTAT----- 7017
 Db 1026 yLysAlaHisGlnLysGlnLysLeuSerGlnProAsnLeuSerSerArgLysAspMetSerAsn 1046
 QY 7018 -----GACGTAAGCACAAAGATGTGACAGAGCGGTGTGAACATATCAAAA 7064
 Db 1046 rProGlnProLeuAspAsnThrSerGlnLysThrThrGlnLys-----IleSerAsnAs 1064
 QY 7065 TTTTATTAAGAGTGAAGAACTGAATATGAAGCAACAAAGAAAGAAAGTTCAAAAGAT-- 7122
 Db 1064 nGlnLysLysValAsnGlnLysGlnLysPheLysGlnLysGlnLysGlnLysGlnLys 1084
 QY 7123 -----AAAGATGGCAAAAGTATAGGATTAATCC 7151
 Db 1084 eValLeuLysSerHisMetAsnArgLysSerAspAspGlnLysLeuThrAspGln----- 1102
 QY 7152 TTCTACTGAAGAGCATAGAGAGCAACATGTGCTCATGAATATTAACATGAATAT 7211
 Db 1103 -----AsnSerAspLysSerThrValAsnAspGlnSerGlnLysAlaGlnLysMet 1120
 QY 7212 AAAAGATTAATATGCAATTAAGATGTGCTGATGCAAAAAACCTTCTTCAACATAC 7271
 Db 1120 Lys-----GlyAsn--AspThrSerGlnLysSerHisAsnSerSerGlnHisIle 1136
 QY 7272 AAAAACAACAACAATCAATCAATCCATGCTAATGATATGCAAGATCCCTGGAT-- 7329
 Db 1136 eGlnSerAspGlnGlnLys-----AsnAspMetLysThrValGlnLysPhe 1151

Alignment Scores:	1.09e-38	1435
Prod. No.:		
Score:	618.00	Matches: 339
Percent Similarity:	36.80%	Conservative: 223

QY	3301	GAAGCTTGTAATGAATTAATGAATTAATAGATTAATGGAAGAACA-----TGC	3351
Db	20	LysAlaArgAsnGluTyr--AspLe-----LysGluAsnGluLysPheLeu	34
QY	3352	AATTATATATATCGATTAATATACAAAGATTTACATGACAAAGCAAAATGCTGTTAGTAAT	3411
Db	35	AspValI TyrLysGluLysPheAsnGluLeuAsnLysPheLysI TyrGlyAsnValGlnLys	54
QY	3412	ACTGGTATTGGAAGCTTCCAGTACTGCCAAAATCATATAGACAGAGATGTATGAAATT	3471
Db	55	Thr-----Lys-----AspLysLysIlePheThrPhe	62
QY	3472	TTG---TCGGAATATATACCAACAAATGCTGGCAAAAGTAAATTAACGT-----	3516
Db	63	IleGluAsnLysLeuAspIleLeuAsnAsnSerLysPheAsnLysArgTrpLysSerTyr	82
QY	3517	GGTACTACTGTAAGAAAGTGCTGTCATTGCGTACTACACACCCAGTATGAAATGTTGGAGCA	3576
Db	83	GlyThrProAspAsn-----IleAspLysAsnMetSerIleLeuAsn-----	96
QY	3577	TATCTCCATGTATACAGAAATTTTGATGANTGTGCAGTACACAAATGATGTTTGTGATGA	3636
Db	97	--LysHisAsn--AsnGlu---GluMetPheAsnAsnAsnTyrGlnSerPheLeuSerHis	114
QY	3637	AAAGTATGATGTAAGATTAACGAAAAATATGCTTTTAGATATAAACCACAGGACCATGAT	3696
Db	114	rSerSerLeuIleLysGlnAsnLysTyrValProIle-----AsnAlaValArgVal	131
QY	3697	GGTCGGTGGTGTTGTAATAAGTGGATCGMAACCAACAGGATACAGATATAAACGAAATAA	3756
Db	131	I SerArgIleLeuSerPheLeuAsp-----Lys-----	139
QY	3757	AAAGCGGAAGAAAGATGACGAAATGTAATAACAGTG-----ATGATATATCTTAA	3806
Db	140	-----SerArgIleAsnAsnGluLysAsnThrSerSerAsnAsnGluValLeuLeu	156
QY	3807	AGAA--AACGATGAGAAAGAACACAGATAGAA--GATTGTCATCCAAAAGCAATAGTAA	3860
Db	156	rAsnCysArgGluLysArgLysGlyMetLysTrpAspCysLysLysLysAsnAspArgSer	176
QY	3861	TGATATATCCGATTTGGCAATCGCGAAATATTAATTAGTGAAGACCTCGTGTGTAT	3920
Db	176	rAsnTyr-----ValCysIle	181
QY	3921	GCCCCCTAGAGACAAAGTTATGCGTACATTTCTTGCCAAATGATTAATGAATAATAAAA	3980
Db	181	eProAspArgArgIleGlnLeuCysIleValAsnLeuSer-----IleIleLys	197
QY	3981	ATATACATACACAAGTTAATTAAAAAGACCTTTCATCAAAATCTGCAGCAGCAAGAAACATT	4040
Db	197	srThrTyrThrLysGluThrMetLysAspHisAspHisIleGlnIleAsnSerLysGlu---	215
QY	4041	CTTCTCATGTGTTATTATTAATAAGTAAAGATGGGTGTAAGGAATGAACTCGATTAAGAATT	4100
Db	216	-----SerGlnLeuLeuLeuLysLysAsnAspAsnLysTyrAsnSer-----	229
QY	4101	AAAAAGAGCAAAATTCCTCCCGCATTTTGTAGATCCATGTTCTACACATTTGGAGATTA	4160
Db	230	-----LysPheCysAsnAspLeuLysAsnSerPheLeuAspTyr	242
QY	4161	TAGAGATTTTTTATTTGGACAGAGATATTCAAAAGCTCATGTGTAGAGGAAGCTAAATGAAA	4220
Db	242	rGlyHisLeuAlaMetGlyAsnAspMetAspPheLys--GlyTyrSerThrLysAlaGln	261
QY	4221	AGAGCAAAATAGATTCTTTTCAAAAATGTGACCAAAAATCTCTAATGGAATAAACA--	4278
Db	261	uAsnLysIleGlnGlnValI PheLysGlyValAlaHisGlyGluIleSerGlnHisIleLysIleTyr	291


```

Db 1136 egluseraspInclInlys-----AsnaspmetylstHvalglAspre 1155
QY 7330 -----TATGTCTCTGACGAATTTAAACAAGTGTGAGTGTCTGAACTTTCAAAA-- 7377
Db 1151 uGLyTtHrHIsValdInAsngInleserValProvalTHrGInlIeAspgLulY 1171
QY 7378 -----AAGGATCTATGATTCCTCTACAAAAAAATATTACTGAACCTAAATACC 7425
Db 1171 sleuargGluSerlyseLysleuLyIleHs-----LysAlaGluGluGluArgLeuSe 1189
QY 7425 TATGAATTTGTGTAGAGAAAGACACATATTTATTTATCTAAAGACAGAAAATATATGCA 7484
Db 1189 rHIsThAspRIleHIsLys-----IleAsnProGluAspArgAsnSerAsnTh 1205
QY 7485 TATTACCTTAAAGAAAATTTTATACCTTATAGCTACACAAAAGAAAAGAAAGTAAAA 7544
Db 1205 rleuHIsleuLyAsp-----IleArgAsnGluGluAsnGluArg 1218
QY 7545 TAGTGTGACTAAATAAATAT 7563
Db 1218 gHIsleuThAsnGlnAsn 1224

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Score: 431.00 Matches: 210
Percent Similarity: 34.568 Conservative: 128
Best Local Similarity: 21.47% Mismatches: 262
Query Match: 2.22% Indels: 378
DB: 2 Gaps: 44

US-10-087-013-1 (1-10628) x US-08-568-459A-6 (1-749)

QY 7342 GAATTTACAACTGTGAGTGCTCTGAACCTTCAAAAAAGGATCTATGATTCATACAAA 7401
|||||
Db 35 GluPhasInlySile-----PheProGluArgSInValGlnIleHisSileSer 50
QY 7402 AAAATTTACGACCTTAATACCTATGATTTGTGTAGAGAAACAGCATTTATTATCT 7461
|||||
Db 51 AsnIle-----PheLys 54
QY 7462 AAAGAAGCAGAAATATATGATATT-----ACCTTGAAGAAATATATACCT 7512
|||
Db 55 GluTyrLysGluSInSInValAspIleIlePheGluTyrLeuSInTyrGluTyr----- 72
QY 7513 ATTGAGCTTACAAAGAAAGAAAGTAAATAGTTGACATAATATATCTTGGCAT 7572
|||||
Db 73 AsnAspPheCysLysGluSInTyrProGluLeuValSerAlaIleLysTyrAsn----- 89
QY 7573 CCTAAGAAACCTTATGCACCTGATTAATATATGAGAAAGAAACCTTGTGAATATGA 7632
|||||
Db 90 LeuLysAlaProSInAlaLysSerProArgIleTyrLysSerLys-----GluHisGlu 107
QY 7633 GAAGAATAATGCTTT-----AGGTAGATTATGAAATGAAATG 7671
|||
Db 108 GluSerSerValPheGluCysLysTyrHisLysSerLysValLysLysLysTyrPAsnCys 127
QY 7672 TACAAAATTCCAAGATTCATACAGAGAAAGAAAGATATGTATGCTCCAGAGAGAA 7731
|||||
Db 128 TyrSerAsnSInLys-----ValThrLysProGluGluValCysGluProArgArgGlu 146
QY 7732 CATATGCTTATGAGATTTAGATGAATTTAA-----ATTGAACACTTAAGAT 7782
|||||
Db 147 GlnLeuCysLeuGluTyrIlePheLeuIleLeuTyrAspGluSInGluGluLysLysAsp 166
QY 7783 AGTAATATCTCTAAATAATGTTCTGTCAGACAGAAATGAGAAATGACATATATA 7842
|||||
Db 167 His-----IleAsnLysAlaIleAsnTyrGluAlaMetHisLeuLys 180
QY 7843 AAAAATCTCACTCAGAGAAAGCGGTGGCAGATCAATCATATGCTACTATGAATAT 7902
|||||
Db 181 GlnLysTyrGluSInAlaGlyLys-----AspLysIleCysAsnAlaIleLeuGlu 197
QY 7903 AGTTGGCTGATCTGGTGATGATGATGAGAAAGATATGTTACCAATTTGGTGTAC 7962
|||||
Db 198 SerTyrAlaAspIleGlySInIleValArgGlyLeuAspValTyrPArg----- 213
QY 7963 TTACCTCCCGTGAATAATA-----AAATTTATTAAGTTTGTGAAATCATATATGA 8013
|||||
Db 214 -----AspIleAsnThrAsnLysLeuSerGluLysPheGlnLysIlePheMet 229
QY 8014 AAATGAGAAATATAAATAAGGTAGAAATATAATACACAGATGTACAAAGCTTGTCT 8073
|||||
Db 230 GlyLysLysSerArgLysGlnAsnAspAsnSInLys-----ArgAsn 245
QY 8074 GCTTGGTGATCTGATTAAGAAAGATATTTGGAAGCATG----- 8115
|||||
Db 246 LysTyrTrpLysGlnArgAsnLeuIleTyrSerSerMetValLysHisIleProLys 265
QY 8116 -----ACGTGCAAGCAGCAGAGATGCAAAACTTTTGAAGAAAGAGATGATGA 8169
|||||
Db 266 GlyLysThrCysLys-----ArgHisAsnAsn 274
QY 8170 TTTGAGCAGATAACATTATACAAAGATAGTGTGACATAGAGACATCCACCTGTTGAT 8229
|||||
Db 275 PheGluLys----- 277
QY 8230 GATTATATACCTCAACGCTTTCGATGATGATGATGATGCTGCTGATATTTATTAAGCA 8289

Db 278 -----IleProGlnPheLeuArgTyrPLeuLysGluTyrPLeuLysGluPheCysGluGlu 295
QY 8290 CTGATGGAATAATTTGAAAAATTTAAATATCATGATCAGCTGTAATACATGACAGCA 8349
|||||
Db 296 MetCylThrGluValLysGlnLeuGluLysIleCysGluSInLysAsnCysSerGluLys 315
QY 8350 TGCAGAGATGATTTATGATGAATAATAGTGTGACAGCTGTAAACAGAGATGTCAAGATAT 8409
|||||
Db 316 -----LysCysLysAsnAlaCysSerSerTyr 324
QY 8410 AAAAATTTGCTTCTTAATGGAATCTCTATGATATACATATCAATTAATACAAAGAA 8469
|||||
Db 325 GluLysTrpLysGlnArgLysAsnGluTyrAsnLeuIleSerLysLysPheAspSer 344
QY 8470 TTGTATGACACACCAATATATATACAAATCTCTACTATGATCATGTTCAAAATTTGTA 8529
|||||
Db 345 -----AspLysLysLeuAsnLysLysAsnSInLysTyrAsn----- 356
QY 8530 CAAAAGTTGAACCTTTAAAGTAAAGTGTCTGTGAGAGCTTTTCTGATATCTTCAT 8589
|||||
Db 357 -----LysPheGluAspSerLysAlaTyrLeuArg 366
QY 8590 GAAACAGTAG-----TGTGATATTAATTTATGAAATATGATGCTTCTTCAATATA 8646
|||||
Db 367 SerGluSerLysGlnCysSerAsnIleGluPheAsnAsp----- 379
QY 8647 CGAATATGCTTTTGAAGAAACACCAAAAGTATTAAGAGCTTGACGT-----TGTACA 8703
|||||
Db 380 GluThrPheThrPhe-----ProSInLysTyrLysGluAlaCysMetValCysGlu 396
QY 8704 CTACCTTCTAAGAT-----CCATGTGATAT-----TGCTTACCGATCAAA 8745
|||||
Db 397 AsnProSerSerSerLysAlaLeuLysProIleLysThrAsnValPheProIleGluGlu 416
QY 8746 AACCAATATGATGATGATGATTAACAACTTTTACCTTCTGCTGGAAGATGATTAATGAT 8805
|||||
Db 417 SerLysLys-----SerGluLeuSerSerLeuThrAspLysSerLysAsn----- 431
QY 8806 AATATCTTGTATTAATGGAACGCATACCTTGTCTTAATGTTACGATGATTAACAAAGT 8865
|||||
Db 432 -----ThrProSInSerSerGlyGly----- 439
QY 8866 GTATGATTCCTCCAGAGAGACATTTATGTACAAAGCCTATCACTGATATATATAT 8925
|||||
Db 440 -----AsnTyr 441
QY 8926 AGAAAGCTGATAAGAAATTTTAAAAA-----AAACTT 8961
|||||
Db 442 -----GlyAspArgGlnIleSerLysArgAspValHisHisAspGlyProLysGlu 459
QY 8962 CTACTCTGCTTCTCACTCAAGCAGCAATTTAGTCAAAAATTAATATGGAAGAGAG 9021
|||||
Db 460 ValLysSerGluGluSInGluValProLysIleAspAlaValLysTyrGluAsnGlu 479
QY 9022 TTGTGC-----TTGAGCAGATGAATAATAGTATAGCATTTATCCGAT 9066
|||||
Db 480 PheThrSerAsnArgAsnAspIleGluGluLysGluLysSerLysGlyAspHisSerSer 499
QY 9067 ATAATTAAGCACTGATATGATGAGACATTTATATCTGAAAAATTAATAAATTTT 9126
|||||
Db 500 ProValHisSerLysAspIle-----LysAsnGlu 509
QY 9127 GAAACATCAATATGAGCAGCAGAAATCGTAAACATGCTGGGAAATAATATAGACGTGAC 9186
|||||
Db 510 GluProGlnArgValValSerGluAsn----- 518
QY 9187 ATATGCAAGCTATGTTATGTGATATATAATGCTACTCAAAAGTAAACATTAAGATGAA 9246
|||||
Db 518 ----- 518
QY 9247 GATGTGTCAATTTACCAAGATGAGAAACATATCAGTTTCTTCTGCTTATTTGAA 9306
|||||

```
Db 519 -----LeuProLysIleGlu----- 523
QY 9307 TGGCAAGCAAGCATGTAAGAAACATGTAAGTTCATTAACAAATGT 9366
Db 524 -----GluLysMetGluSerSerAspSerIleProIle----- 534
QY 9367 CCTCGTTCAAGCAAGATTAATTTTGAGCGTCAGAAATTTTAAGCAACCTGGATGTCAG 9426
Db 535 -----ThrHisIleGluIleGlu----- 540
QY 9427 AATGATATTAGAAAATATATTAGCTTGAAATATATTGATAAAAAATACATGCAAAATCTA 9486
Db 540 ----- 540
QY 9487 AATATAAATATTAAGCAATTAAGATCAATCTTCAGTAATATAGACAAATTAACCATCT 9546
Db 541 -----LysGluGlnSerSerAsnSerSerAspAsnProAla 553
QY 9547 -----GAGAA 9552
QY 554 ValValSerGlyArgGluSerLysAspValAsnLeuHisThrSerGluArgIleLysGlu 573
QY 9553 AATGTCAGTCATATATATAAATCAAAAGATTCCTCAATGCCCTTGAGTGA----- 9603
Db 574 AsnGluGluGlyValIleLysThrAspSerSerLysSerIleGluIleSerLysIle 593
QY 9604 -----AATCATATAATGAATAGTAATAGTACAGAAACAAATAATAT 9642
Db 594 ProSerAspGlnAsnAsnHisSerAspLeuSerGlnAsnAlaAsnGluAspSerAsnGln 613
QY 9643 GAAATTAATGAA-----TTCAAGAAGTACTTAAAAATTTATATCT 9684
Db 614 GlnAsnLysGluThrIleAsnProSerThrGluLysAsnLeuLysGluIleHis--- 632
QY 9685 GATTATATTTTGTGAGATGCAACACAAATAATCATGTACTAGATGGAATATAAA 9744
Db 633 -----TyrLysThrSerAspSerAspHisGlySer 643
QY 9745 GAAGAAGCAAAACAGTTCGTCCTTAAGCAGCTATTTCTTTACACCCCATGTAGATTCT 9804
Db 644 LysIleLysSerGluIleGluProLysGluLeuThrGluLysProLeuThrAspLys 663
QY 9805 TTCTATCAAGCACCTTTATTCACACATGAGTACACAAATATGATCTAAAAATGAT 9864
Db 664 -----LysThrGluSerAlaAlaIleGlyAspLysAsnHisGlu 676
QY 9865 ATATTGAAAGTAGT----- 9879
Db 677 SerValLysSerAlaAspIlePheGlnSerGluIleHisAsnSerAspAsnArgAspArg 696
QY 9880 ---ATCTCTGTTGTTATGTA-----TCGCGCTTAGCTTTGATACCGCTT 9921
Db 697 IleValSerGluSerValValGlnAspSerSerGlySerSerMetSerThrGluSerIle 716
QY 9922 CATTTCATGAGAAAAAATTCAAATCGTCTGCGACTTGCTGCTACTGAAT 9975
Db 717 ArgThrAspAsnLysAspPheLysThrSerGluAspIleAlaProSerIleAsn 734
```

Search completed: April 28, 2003, 12:04:25
Job time : 577.5 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 11:01:12 ; Search time 653.5 Seconds
(without alignments)
6701.970 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455
Sequence: 1 atgggggtctctgcgaata.....atatatgatgaagtatg 10628

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Archived: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_n2p_model -DEV=xlp
-Q/csn2.1/uspro.spool/US10087013/runat.28042003.102845.20267/3pp.query.fasta_1.10823
-DB=SPRMBL.21 -OEFW-fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=Diosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=plc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10087013.eccn.1.1115_etunat.28042003.102845.20267 -NCPV=6 -ICPV=3
-NO_XIPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV TIMEOUT=120
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPRMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	
1	19407	99.8	3542	5	Q905M2	Q905M2 plasmodium

2	12954.5	66.6	3287	5	Q8T326	Q8T326 plasmodium
3	7367	37.9	1615	5	Q8T325	Q8T325 plasmodium
4	4515.5	23.2	2706	5	Q15870	Q15870 plasmodium
5	4260	21.9	3006	5	Q26032	Q26032 plasmodium
6	4010	20.6	3026	5	Q26030	Q26030 plasmodium
7	3501	18.0	3078	5	Q26031	Q26031 plasmodium
8	3328.5	17.1	2527	5	Q95W83	Q95W83 plasmodium
9	3154	16.2	2924	5	Q25733	Q25733 plasmodium
10	2932	15.1	1685	5	Q9U4A2	Q9U4A2 plasmodium
11	2869	14.7	2658	5	Q8T5G0	Q8T5G0 plasmodium
12	2817	14.5	510	5	Q8T6L0	Q8T6L0 plasmodium
13	2727	14.0	2647	5	Q90580	Q90580 plasmodium
14	2726.5	14.0	2664	5	Q26033	Q26033 plasmodium
15	2678.5	13.8	2212	5	Q94657	Q94657 plasmodium
16	2556.5	13.1	494	5	Q8T6K7	Q8T6K7 plasmodium
17	2480.5	12.7	492	5	Q8T6K9	Q8T6K9 plasmodium
18	2457	12.6	2710	5	Q9XZB8	Q9XZB8 plasmodium
19	2449.5	12.6	494	5	Q8T6K0	Q8T6K0 plasmodium
20	2441.5	12.5	465	5	Q8T6K4	Q8T6K4 plasmodium
21	2413.5	12.4	494	5	Q8T6K5	Q8T6K5 plasmodium
22	2412.5	12.4	492	5	Q8T6K6	Q8T6K6 plasmodium
23	2403.5	12.4	496	5	Q8T6K8	Q8T6K8 plasmodium
24	2335.5	12.0	461	5	Q8T6K3	Q8T6K3 plasmodium
25	2303	11.8	2135	5	Q61077	Q61077 plasmodium
26	2258.5	11.6	460	5	Q8T6K1	Q8T6K1 plasmodium
27	2076.5	10.7	1327	5	Q9NFB4	Q9NFB4 plasmodium
28	2055.5	10.6	427	5	Q8T6K2	Q8T6K2 plasmodium
29	1890.5	9.7	2169	5	Q97312	Q97312 plasmodium
30	1773	9.1	2197	5	Q96296	Q96296 plasmodium
31	1755	9.0	2163	5	Q9NFB6	Q9NFB6 plasmodium
32	1755	9.0	2182	5	Q26034	Q26034 plasmodium
33	1744	9.0	2209	5	Q97324	Q97324 plasmodium
34	1731.5	8.9	2228	5	Q60991	Q60991 plasmodium
35	1709.5	8.8	2277	5	Q9U0G5	Q9U0G5 plasmodium
36	1614	8.3	2209	5	Q9U0G6	Q9U0G6 plasmodium
37	1586	8.2	1711	5	Q96108	Q96108 plasmodium
38	1542.5	7.9	2042	5	Q25766	Q25766 plasmodium
39	1249	6.4	1729	5	Q25734	Q25734 plasmodium
40	924	4.7	173	5	Q96735	Q96735 plasmodium
41	835.5	4.3	431	5	Q96294	Q96294 plasmodium
42	750.5	3.9	2647	5	Q9U4X0	Q9U4X0 plasmodium
43	665.5	3.4	438	5	Q9Y1N7	Q9Y1N7 plasmodium
44	658.5	3.4	455	5	Q9Y1N6	Q9Y1N6 plasmodium
45	651	3.3	118	5	Q964S0	Q964S0 plasmodium

ALIGNMENTS

RESULT 1
ID Q905M2 PRELIMINARY; PRT; 3542 AA.
AC Q905M2;
RX 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAY-2002 (TREMBLrel. 20, Last annotation update)
GN FC3 CSA ligand (Fragment).
OS VAR.
OC Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=20006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheidig C., Baruch B., Oishi S., Fujii N.,
Fusai T., Parzy D., Miller L.H., Gysin J., Scherf A.,
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
sulfate A: A receptor for human placental infection.",
RL EMBL; AJ13811; CAB59840.1; -
DR InterPro: IPR001219; Neurotoxin.
DR InterPro: IPR004258; PFEKP.
DR Pfam: PF03011; PFEKP; 1.

DR PRINTS: PR00284; TOXIN.
 FT NON TER 3542 3542
 SQ SEQUENCE 3542 AA; 413089 MM; 970D85EE8BD2A2EC2 CRC64;

Alignment Scores:

Pred. No.:	0	Length:	3542
Score:	19407.00	Matches:	3542
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	99.75%	Indels:	0
	5	Gaps:	0

US-10-087-013-1 (1-10628) x Q9U5M2 (1-3542)

```

QY 1 ATGGGGTCTCTCGCAATATTTATTTATTTAAATGGGGAATGCACATCATCTTAAG 60
   |||||
Db 1 MetGlyPheSerCysLysTyrPheIleIleLysMetGlyAsnAlaIleSerLeuGlu 20
QY 61 CGAGATGCTAAAGCCCTATTATTAAGAAAGTCACAAAAGTCAGAGAATGTTTGAA 120
   |||||
Db 21 GlyAspAlaLysSerProIleIleLysGlnSerIleLysSerAlaIleGlnValLeuGlu 40
QY 121 CGTATGGCAAAATATTAAGCATCCATCAAAATATGCAAAAGACATGTGATTCGTTG 180
   |||||
Db 41 ArgTyrAlaLysAsnIleArgHisProSerLysTyrAlaLysGlnHisValAspSerLeu 60
QY 181 AAAGGGATTTGAGCAAGAGCAATTCGTGGTGCTCTTACGCCAGTAATTAAGCAT 240
   |||||
Db 61 LysGlnLysPheLysLysLysAlaGlnPheArgGlyGlyProSerThrProValAsnLysHis 80
QY 241 AATTATATATATTCATATTCATGTAATTAATTAAGATCAAGACATCAATTAATTCGATAT 300
   |||||
Db 81 AsnTyrTyrTyrProTyrProCysAsnLeuAspHisLysGlnHisThrAsnLeuArgTyr 100
QY 301 GATGATGTGATTTGAGACATCCTTGCCATGGTAGAGCAAAACCGATTTGATGAAGAT 360
   |||||
Db 101 AspAspAlaLysLeuArgHisProCysHisGlyArgGlnAsnArgPheAspGluAsp 120
QY 361 GAAGAATCTGAATGTGAAATTAATACGTAAATTAATTAAGAAATTAATGATGCTATGACC 420
   |||||
Db 121 GlnGlnSerGlnCysGlnLysAsnLysIleArgAsnTyrLysArgLysAsnAspAlaIleHis 140
QY 421 TGTGCCACCTTGAGAGAGACATATGCTGATTAATAAACTTGAAGCTTAATGATATA 480
   |||||
Db 141 CysAlaProProArgArgArgHisPheCysAspLysAsnLeuGlnAlaLeuAsnAspIle 160
QY 481 AATACCAAAATATTCATGATTAATGGAATGTACTAGTACAGCAAAATATACAGAGT 540
   |||||
Db 161 AsnThrGlnAsnIleHisAspLeuGlnLysAsnValLeuValThrAlaLysTyrGlnGly 180
QY 541 GAATCAATGTTAATATCATCATCACTTAAGAACTTCAAGCGCTTGATGCTCTTCA 600
   |||||
Db 181 GluSerIleValAsnAsnHisProHisLysGlyThrSerAspAlaCysThrAlaLeuAla 200
QY 601 CGAAGTTTGCAGATATAGTGATATTTGAAGAGAAATAGATATGTTTAAACCAAGTGC 660
   |||||
Db 201 ArgSerPheAlaAspIleGlyAspIleValArgGlyIleAspPheLysProAsnVal 220
QY 661 CATGCAAGTAGAAGACGGCTCTCCAGAGGCTTTCAAGAAATATACATGATGAAGAA 720
   |||||
Db 221 HisAspLysValGlnThrGlyLeuArgGlnValPheLysLysIleHisAspGlyMetGln 240
QY 721 GATCAAGTAAGAAATGATTAATCAATCCATGATGATCTGGAATTTATTAATTAAGAA 780
   |||||
Db 241 AspGlnValLysAsnAspTyrAsnProAspLysSerGlnLysAsnTyrLysLeuArgGln 260
QY 781 GCATGTGGGAATGTGAATAGAAATAAAGTATGGAACCTTAACATGATGATGATCAT 840
   |||||
Db 261 AlaTyrPheAsnValAsnArgAsnLysValThrGlnAlaIleLysHisCysAspAlaSerTyr 280
QY 841 AAATCTGATATTTATTAAGCAATGAGAAATATACCATTAATTTTCAATCTTAATGC 900
   |||||
Db 281 LysSerGlnTyrPheMetGlnSerGlnSerAsnThrProLeuPheSerAsnProLysCys 300
  
```

```

QY 901 GCCCATTAACAAGAAAGTTCTCTACCAATTTAGATTATGCTCCCTCATATTTAGCTGG 960
   |||||
Db 301 GlyHisLysGlnGlnLysValProThrAsnLeuAspTyrValProGlnTyrLeuArgTyr 320
QY 961 TTCAGCAATGGGAGAGAGCTTTTCCGAAAAAGAAATATTAATTTGAAAAAGTCAAG 1020
   |||||
Db 321 PheAspGlnTyrProLysGlnGlnPheCysArgLysArgAsnIleLysLeuLysValLys 340
QY 1021 GACTCTGTGTAATGACAAAGAACGCTTATATTTAGTCAATGATGATGATGATGATG 1080
   |||||
Db 341 AspSerCysArgAsnAspLysGlnArgLeuTyrCysSerHisAsnGlnHisAspCysThr 360
QY 1081 ACACATTTTGGAAAAAGGATTTTGGCATTTGGCATTAATAGTACTGATGCTGCACT 1140
   |||||
Db 361 ThrThrIleTyrPheLysGlnLysLeuHisLeuAspAsnLysCysThrAspCysSerThr 380
QY 1141 AAATGCAAGTTTGTGAAGTTGTTAGGGAATCAACAAGACATTTAAAAACAAAA 1200
   |||||
Db 381 LysCysLysValPheGlnValThrPheLysGlnGlnGlnGlnAlaPheLysLysGlnLys 400
QY 1201 GAAAAATATGAAAAAATATACATCATATTTATGCAAGATACCAATTTGCAATTAAT 1260
   |||||
Db 401 GlnLysTyrGlnLysGlnIleGlnSerTyrLeuSerAsnAspAsnLysPheValAsnAsn 420
QY 1261 AATTATAGTGAATATTTATTAACAAATTTTATTAAGAAATTAAGAAAGCAATATGCAACT 1320
   |||||
Db 421 IleAsnSerGlnTyrTyrLysGlnPheTyrGlnLysLeuLysGlnThrGlnTyrAlaThr 440
QY 1321 AATGCACTTTTATTAATTTACTAAATGAAGAAAGATTTTAAGAGAGATTTACAGGA 1380
   |||||
Db 441 AsnAspThrPheLeuAsnLeuLeuAsnGlnGlnLysTyrCysLysGlnGlyLeuProGly 460
QY 1381 GAAAAATATTTACTTTTACTAACAAGCTGATGATACAAAGGATATTTATTCGTTGAA 1440
   |||||
Db 461 GluLysAspIleThrPheThrAsnSerAlaAspAspLysGlyIlePheTyrArgSerGln 480
QY 1441 TATTGCCAAGTGTCTCCGAGCTGGGGGTCAAAATGTATGATGTTAAATACACACAAA 1500
   |||||
Db 481 TyrCysGlnValCysProAspCysGlyValLysCysAspLysIleLysTyrThrHisLys 500
QY 1501 TCAGATATGATGCTGAAGCTGTAATTAATTAAGAACTATTAACCTCAGTGGGTGGAAG 1560
   |||||
Db 501 SerAspAsnAspArgGlnArgValAsnAsnGlnAspTyrLysProProThrGlyValLys 520
QY 1561 CTAATAATATCACTGCTCTTATAGTGATGATGATGATGATGATGATGATGATGATG 1620
   |||||
Db 521 ProThrAsnIleThrValLeuTyrSerGlnLysGlnGlnLysAspIleThrGlnLysLeu 540
QY 1621 GAAAAATTTTGTAAACAGCTCACTAATTTCAAAAGATAAAAATTAACAAAATGGAATGC 1680
   |||||
Db 541 GluAsnPheCysAsnSerSerThrAsnTyrLysAspLysAsnAsnGlnLysThrProLys 560
QY 1681 TATTATAGGATGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1740
   |||||
Db 561 TyrTyrLysAspLysAsnLysHisAsnArgCysLysLeuGlnGlnAsnThrGlnIleAsnAsn 580
QY 1741 GATTAATCTTAATTAATTAATTAATTAATTTTGAATTAATGATGATGATGATGATGAT 1800
   |||||
Db 581 AspAspProLysIleIleSerPheHisAsnPhePheGlnLeuThrValThrTyrLeuLeu 600
QY 1801 AGGGAATATTAAGGGAATGACAAACTTAATTTGATTAATTAATTAATTAATTAATTAAT 1860
   |||||
Db 601 ArgAspThrIleLysTyrAsnAspLysLysTyrCysIleAsnAsnThrThrHis 620
QY 1861 TGTATTTGATGAATGTAAACGAATGCTTATGTTTGAAGATGCTTAAACAAAGAA 1920
   |||||
Db 621 CysIleAspGlnCysAsnArgAsnGlnCysLeuCysPheAspArgThrValLysGlnLysGln 640
QY 1921 GAAGAATGGAATAGTATTAAGAACTGTTCACAAAAAAAGAAATATTAACGAATGCTAT 1980
   |||||
Db 641 GlnGlnTyrPheSerIleLysLysLeuPheThrLysLysAsnIleGlnGlnSerTyr 660
  
```


QY	1981	TATGTAATTAATTATATATCTTTTGAAGGTATTTTAAAAAGTATGCAATTAACCTGAC	204
Db	661	TTTSTSerASnTlLeASnSneUpneGUGUyTYrPhnePheUySValMeTASpLysLeuAsp	680
QY	2041	AAAGATCAACCAAAATGGAAGACATTAATGAGAAATATTAAGCAAAATTAAGAGCTTT	2100
Db	681	LysAspLunAlaLysTrpLysGluUeMetGluAsnTlLeYsArGlyLysAsnGluPhe	700
QY	2101	TCCAAATTTGGAAAAATTAATAGGACATATTTAGAGAAATGCAATAGAACCTTTGTAGATCAC	2160
Db	701	SerASnLeuGluASnSnsnArGAspTYrLeuGluASnAlaTlLeGluLeuLeuASpHis	720
QY	2161	TTTAAAGAAACCTCCACGATATCTATAAGCATATATACAAACGAAGCATGTGAAACATCC	2220
Db	721	LeuYsGluThrAlaThrTlLeCYsLysAspAsnAnThrASnGluAlaCYsGluThrSer	740
QY	2221	CATATATCAACAACAACCCGTCGTGTAAACCTCTGAGGACGACCAACCATTAATAAT	2280
Db	741	HisASnAlaThrTrhASnProCYsValLysProAlaGlyGlyThGlnProThLysAsn	760
Db	2281	ATAAAAGAAATAGCACAAATCTTTTAAAGAGAGTGCATGACGAGGAAGACGAAATCGTGT	2340
QY	2341	CTTCATTAATTTGAAAGGAAGGACACACGAAGCATATATATTAACGTGGGGGTAGAGAAAG	2400
Db	781	LeuThLysLysLeuYsGlyLysAlaThASnGUGUyTlLeTYrLysArGlyGlyLArGArGlys	800
QY	2401	GACTTCAGAGACAAATTTATGTAGATAATATGATTAAGATAACATCTTAATCGATATCTTGTTTT	2460
Db	801	AspPheLysAspHisnLeuCYsArGtlLeMetTlLeYsHisSerASnArGAsnLeuGlyPhe	820
QY	2461	TCAAATGACCATGTGATGGCAAGGACACAGGTGATGTATACAAACAAGTTTGTGCTA	2520
Db	821	SerASnGlyProCYsAspGlyLysGlyThrGlyLysArGlyTlLeGlnThrArGpHeValAla	840
QY	2521	GGAACGTAAATGGGAAGACTGGATCCGGAACACATGCGCTAAAGATCACGAAGATTTATATG	2580
Db	841	GlyThGtlUwTrpGluValAspProGluHisMetArGlyAspHisGluAspValTlLeMet	860
QY	2581	CCCTCCATGAAGACGACATATATATACATCCAAATTTGGAACATTTTCAAAACGATGATCAC	2640
Db	861	ProProArGArGArGhIStLeCYsThrSerASnLeuGluHisLeuGlnThrAspAspHis	880
QY	2641	CCACTTAATGTAATATGTGTGATGATTAATTAATTCCTTTTGGGGATGTCCTT	2700
Db	881	ProLeuASnGlyASnTlLeValAspAspLeuValASnASnSerPheLeuGlyAspAlaLeu	900
QY	2701	CTATACGCAAAATATGACGAACAACAAGATATATACAAATGTATTAAGAAAGAAATACCTA	2760
Db	901	LeuSerTlLeYsTYrGluAlaASnLysTlLeTlLeArGMeTYrLysGluLysASnASnLeu	920
QY	2761	AAGGGCCCCAAAGAGTACTGACCCCAAAACACGACACATGATGTGCGAGCTATACGT	2820
Db	921	LysGlyProLysGluValThrASpProLysHisGlnThrThTlLeCYsArGAlaTlLeArg	940
QY	2821	TACAGTTTTCAGATATAGGTGATATATTAATTCGAGAGAGATCTCTGGAAGAAACGCT	2880
Db	941	TYsSerPheAlaAspTlLeGlyAspTlLeArGValArGAspLeuTrpGluArGASnGly	960
QY	2881	GACATGTTAAAGTCGACAGACATTTGGAACCTGTTTTGGTAAATATACATTAAGTCACTC	2940
Db	961	AspMetValLysLeuGlnGlyHisLeuGluThrValPheGlyASnTlLeHisLysSerLeu	980
QY	2941	AAAGGCAAAAGGAATGTAAATTAATTAATGATGATGGCCCCAAATATTAAATTAAGGGAA	3000
Db	981	LysGlyLysLysASnAspLysTrpYsASnAspAspAlaProLysTYrLeuLysLeuArGlu	1000
QY	3001	AATTTGTGGGAACCTATATAGAGCCCAATGTATGGGAAGCCATGAATTCGATATAAATAT	3060
Db	1001	AsnTrpTrpGluAlaASnArGAlaLysValTTrpGluAlaMeTlLeCYsAspTlLeYsTr	1020
QY	3061	TTGAGAGATTAATTCGGGACCAACATCACACAACAAGTACTTATTCGGGATATAGTATCAT	3120

[illegible]

Db 1381 PheTyrThrPheGlyAspTyrThrAspPheLeuPheGlyThrAspIleSerLysGluHis 1400
QY 4201 GGTGAGGGAATGTAACATAAGAGCAAAATAGATCTCTTTTCAAAATAGTGACCAAAAA 4260
Db 1401 GlyGluGlySerLysLeuLysGluGlnIleAspSerLeuPheLysAsnGlyAspGlnLys 1420
QY 4261 TCTCCATAATGAAAAACGCCAAGATAGGTGACAGAACATAGTCATAGATATGGGA 4320
Db 1421 SerProAsnGlyLysThrArgGlnIleTyrThrGlnHisSerHisGlnIleTyrGln 1440
QY 4321 GCTATGCTATGTGCTAGTAAATGAGGCAAAAAAGATGATTTACCGCAAACTAC 4380
Db 1441 AlamelLeuCysAlaLeuValLysIleGlyAlaLysLysAspSerPheThrGlnAsnTyr 1460
QY 4381 GGTTCACAACAGCTCAAAATTTAGTACAAAAGCACCACTTGGAGGATTTGCCAAA 4440
Db 1461 GlyTyrAsnAsnValLysPheSerAspLysSerThrThrLeuGluGlnPheAlaLysArg 1480
QY 4441 CCCCAGTTTTCAGTATGCTAGCAACGAATGTAGCGACGCTATGCTATACCGCAAAA 4500
Db 1481 ProGlnPheLeuArgTyrPheLeuThrGlnTyrThrAspSerPyrCysTyrThrArgGlnLys 1500
QY 4501 TATTTGAAGGATGTGCGAGCAAAAAATGTAAGTCAAAATGACCAATGAGGTGATACAG 4560
Db 1501 TyrLeuLysAspValGlnGlnLysCysLysSerAsnAspGlnLeuLysCysAspThrGln 1520
QY 4561 TGTATAAGAAATGCGAGGACTAGCTTAATATATGAAAAAAGAGTGATGCCA 4620
Db 1521 CysAsnLysLysCysGlnAspTyrValLysTyrMetLysLysLysGlnTyrIlePro 1540
QY 4621 CAAGATTAATTTACAAAGATGAAACGCGACAAAAAAGATTCGATAGCAACATTTGGT 4680
Db 1541 GlnAspLysTyrTyrLysAspGlnArgAspLysLysArgPheAsnGlnHisIleGly 1560
QY 4681 GTATGGTTCAGACTTACTGAGACGATGCAACAGATTTACTGTGAACAGAAATTTACT 4740
Db 1561 ValMetValThrAspTyrThrGlnTyrThrAsnAlaThrAspTyrLeuAsnArgLysPheThr 1580
QY 4741 GCTAGTGTGGTATAGAGCTGGAAGTGCCTGTGCTGTCAAAGAAATATACATTTGTA 4800
Db 1581 AlaserCysGlnAspLysProGlnLysSerAlaSerValValGlnTyrAsnIleGlnLeuLeu 1600
QY 4801 GAAAAACAGGCTTACTATGATGCGCAAAACATTTGGGTGCACAAAATTTATTTGAAAT 4860
Db 1601 GlnLysGlnAlaLysTyrThrAspAlaAspLysHisCysGlnCysThrLysPheIleGlnAsn 1620
QY 4861 GACGACAAATATCTAATCAATTTGACTTAAGATTAAGTCCAAAGGATTTAGTAAAGAGCA 4920
Db 1621 AspAspLysTyrThrAsnIleSerLysAspLysCysLysGlyLeuValLysGlnAla 1640
QY 4921 AACACAGGTGCTATTAAGTGGCAAAACAAAGTCTTAATTAATCAATTAATTTGAAAGA 4980
Db 1641 AsnThrGlyAlaIleLysTyrPheGlnAsnLysGlyProAsnAsnTyrAsnAsnLeuLysGln 1660
QY 4981 TTGACTGAGAGTGTGCTTTTCCCTTCGCTGCTGACTACGATATGTTTCATGCAATGGAT 5040
Db 1661 LeuThrGlnLysValLeuPheProSerThrArgLeuArgIleCysPheHisIleLeuAsp 1680
QY 5041 GCGCAATTAATACGATCCAGAACTTAAGATGAAATGGGTGCGCAAAAGATTTGATGGA 5100
Db 1681 GlyAsnTyrThrAspProGlnValLysAspGlnAsnGlyLeuAlaGlyArgLeuMetGln 1700
QY 5101 GTGGGCGGACGAGGCTACAAATTTGGGTCAATTAATCAAAAGAAAAAGAAAAAGAG 5160
Db 1701 ValAlaAlaThrGlnGlyTyrAsnLeuGlnIleTyrTyrLysGlnLysGlnLysGln 1720
QY 5161 AAAATTAAGAGCTGCGATGCGACAAATATCTTAATGAGGTCGCCCTGTGATGCTATG 5220
Db 1721 LysIleLysThrSerAspAlaHisLysTyrSerTyrGlnValProProCysSerAlaMet 1740
QY 5221 AAATTAATGCTTTTATGATTAAGAGATATAATCTAGTATGATTAATTTGGAAGATGAA 5280
Db 1741 LysTyrSerPheTyrAspLeuArgAspIleIleLeuGlnIleAspAsnLeuGlnAspGln 1760

QY 5281 AAACAAAACCGAGAAAAATTTGAGAAAAATTTTAAACAAAAATGAAATCATGTTGGC 5340
Db 1761 LysGlnLysThrGlnGlnLysLeuLysLysIlePheAsnLysAsnGlyThrSerValGly 1780
QY 5341 AAAGGAATGATAGTCTACAGAAATCCCGGTAGTACTGCGGAAAAATTTTCTGGAAC 5400
Db 1781 LysGlySerAspSerThrThrGlnAsnProGlnLysSerThrAlaArgLysPheThrPheAsn 1800
QY 5401 GAAAAATGAAGAAATGTGTGCAACGCAATGATATGCGGGTACAAACGCTGTAGGATGAT 5460
Db 1801 GlnAsnLysGlnLysValIleThrPheAlaMetIleCysGlyTyrLysArgGlyThrAspAsp 1820
QY 5461 GGAATATGAGAAATATGTCAGAAAGTGAAGATCTAAAAAATGTTGTTCTGTACT 5520
Db 1821 GlyAsnSerCylAsnSerAlaArgSerAspGlnAspLeuLysCysGlySerValPro 1840
QY 5521 TCAGATGATGATATATCTTATGGGAAAAATGGCGATGAAAGTATCTGCTATCAGTTTCTT 5580
Db 1841 SerAspAspAspTyrProMetCylLysAsnArgAspGlnGlyThrAlaTyrGlnPheLeu 1860
QY 5581 CGATGCTTTGCCGAATGGGTGAAGATTTTTCAAAACATTAAGAAAAAGAAATGGAGAA 5640
Db 1861 ArgTyrPheAlaGlnTyrPheLysPheCysLysHisLysGlnLysGlnLeuGlnLys 1880
QY 5641 TTGGTGGGGCGGTAAATGATTAATCTGCTGCTATTAAGATTAAGAAAAAGAAATGT 5700
Db 1881 LeuValGlyAlaIleCysAsnAspTyrThrCysGlnAspAsnGlnAspLysArgLysLysCys 1900
QY 5701 ACAGATGCGGTACCAATATAAAAAATTTATGATGATGAGTGAACCACTATGAATAA 5760
Db 1901 ThrAspAlaCysThrGlnTyrLysLysPheIleSerGlnTyrPheProGlnTyrGlnLys 1920
QY 5761 CAAATCAAAAAATATAGTGAAGATTAAGCAAAATATTTCCAGACATCTGTGCGCAAAA 5820
Db 1921 GlnIleLysLysTyrGlyLysAsnLysAspLysIleTyrSerGlnHisProValAlaLys 1940
QY 5821 GATGAGAGAGCGTCCGGAATATTTAGCAAAACAAATTAATAAAAAATTTGAAAAATTA 5880
Db 1941 AspAlaGlnAspAlaArgGlnTyrLeuAspLysGlnLeuLysLysIleCysGlnAsnLys 1960
QY 5881 AGTGAAGATTTGAAATTAATGATGATGAAGATGTCACACAGCATTAATGATGATG 5940
Db 1961 SerGlnAspCysGlnTyrLysCysMetLysAspValSerThrGlnArgLeuThrAspGly 1980
QY 5941 AATAGTCAAAATATGCGCGATCTTAGACGATGAACCAAAAGAGTTGAAGAAAGTGT 6000
Db 1981 AsnSerGlnAsnMetProAlaSerLeuAspAspGlnProLysGlnValGlnGlyLysCys 2000
QY 6001 AATGTCAAGTCCACAGAGGTCCACACAGTGTACGAAGGAAACACCGTCACACAGGGTA 6060
Db 2001 AsnLysGlnValProArgGlyProProArgValArgArgLysThrProSerProAlaVal 2020
QY 6061 TCACTGATATCAAAAGGAGGATGTAAGAAAAAGAAAGCAAGCAAGCGCGCTACAAA 6120
Db 2021 SerLeuIleSerLysAlaThrIleAlaSerLysLysGlnAlaLysThrAlaProProThrLys 2040
QY 6121 CAGCGCAAAAAAGTGAATAATCTTAACAACAAGAAATGCGACCAACAGACCGGACGA 6180
Db 2041 GlnProLysLysValGlnAsnLeuThrThrGlnMetArgAlaGlnThrArgThrArgArg 2060
QY 6181 GCAGCAACAAACAGAAACAGAAACATCAACAGCAACAGCAACAGAAATCTGACGTGGGC 6240
Db 2061 AlaIleGlnGlnThrArgLysArgThrSerThrAlaThrThrThrGlnSerAspAlaGly 2080
QY 6241 ACAATGTAAGAGCCATTTCTTCGAATTAACAGATAGAGGGGTGGAATAGAGGTTGT 6300
Db 2081 ThrMetValLysAlaIleLeuSerAsnLysProAspSerArgGlyGlyIleGlnGlnLys 2100
QY 6301 AATCAAAAAAGTATGAGCAATATCTTAATGAGGGTGTATGTAGTAAAGCTAAAGAA 6360
Db 2101 AsnProLysThrTyrGlnTyrProLysTyrPheLysLysIleValGlyLysSerLysGln 2120

QY	6361	AAATGAAAATGGCATATGCTATGCTCCTACGACAGAAAAAATTTATGTATTAATATATACAA	6420
Db	2121	AsnGlnuSnngLyIleCysMeIProPoaAgaGlyLysLeuCySLIeaSnInIleGln	2140
QY	6421	TATTTAATATGTAAGACTGAAATTAAGCGCTGACATGATATATTAAGAGCGCTTATATTA	6480
Db	2141	TyrIleuSnInrYrGluThrGlnuSnInLysAlaGlySerAspInIleLysGlnAlaPheIleLys	2160
QY	6481	TGTGACCAATAGAAACCTCAATTTTGTGGTTAAATATATATATGAAATCTCGACAGA	6540
Db	2161	CysAlaAlaIleGluThrGlnPheLeuTrpLeuLysTyrIleIleGlnProAlaIa	2180
QY	6541	GAAATGAAATGCAAAATGACAAATCCAGATGCAATGAAATTTAAAGATATATGATATACA	6600
Db	2181	GlnuSnInLeuGlnuSnngLyThrIleProAserGluPheLysArgIleMeTyrTyrThr	2200
QY	6601	TATGTAATATATAAGATATGTTTTTGGACGTATTTCTCAATGATATTAAGTAAAAATATA	6660
Db	2201	TyrIlyAspIlyrYrLysAspMetPhePheGlyThrAspIleSerAsnAspLysLysIleIle	2220
QY	6661	ACTGTACCAATATAGTGTACAAACCATCTTCATGAAATATATAGAAAAACAGATATA	6720
Db	2221	ThrValThrAsnSerValThrThrIleLeuAsnGlnuAsnLysLysLysGlnAspLys	2240
QY	6721	AAAAAAGATGACAAATTTACGTAAATATTTTGGCGAAAAATAAAAATTTATTTGGGAA	6780
Db	2241	LysLysAspGlnGlnLeuAlaGlySLIlePheTrpGlnLysAsnLysLysPheIleTrpGln	2260
QY	6781	GGATGATATATGGAATTACTTTCATCTCACAGACGAAACGAAAAAGAAAAAATATAGA	6840
Db	2261	GlyMetIleYrGlyLeuThrYrHisLeuThrAspGlnuAsnGlnLysGlnLysIleArg	2280
QY	6841	GATATATACAGTACATGACATGACAGACAAACGAGCGCTTCCCTGGAAGATTTGTAAA	6900
Db	2281	AspAsnTrpGlnTrpAsnAspMetThrLysLeuThrProSerLeuGlnGluPheValLys	2300
QY	6901	AGCGCCCAATTTTGGAGATGGTTCACAGATGGCGAGAAATTTTGTATTAAGAGAG	6960
Db	2301	ArgProGlnPheLeuAlaYrPheThrGlnTrpAlaGlnGluPheCysAsnLysLysArgLys	2320
QY	6961	GAAACGTGTTAAAAATTGGAGGGCGGCGCTGAAGAAATATGAGTAAATGATATGATATGAC	7020
Db	2321	GlnGlnLeuLeuLysLeuGlnAlaGlyCysLysGlnYrGlnCysAsnGlySerAsnAsp	2340
QY	7021	GGTAAAGACACAGATATGTCGACAGCGGCTGTACATATCAAAATTTTATTAAGAACTGG	7080
Db	2341	GlyLysThrGlnGlnCysAlaGlnAlaCysValThrYrGlnAsnPheIleLysLysTrp	2360
QY	7081	AAAACTGAATATGAACACAAAGAGAAAGTTCAAAAAAGATTAAGATGGCAAAAGTAT	7140
Db	2361	LysThrGlnTrpGlnArgGlnArgGlnLysPheLysAspLysArgGlyLysLysTrp	2380
QY	7141	AAGGATTAATCCTTCTACTGAAAGACATAGAGAAAGGCAACATGTGCTCATGATATTTA	7200
Db	2381	LysAspIlyrProSerThrGlnArgAspIleGlnLysAlaThrCysAlaInIleGlnTyrLeu	2400
QY	7201	AACATGAAATTTAAAGATATGTCGCATTAAGAAATGTCTTGATGCAAAAAACCTCT	7260
Db	2401	AsnMetLysLeuLysGlnLeuCysGlnLysAsnLysAspCysSerCysMeGlnLysProSer	2420
QY	7261	TCACACTACCCAAAAACACACACACATCACATCATCCGATGCTAATGATATGCGAGAA	7320
Db	2421	SerGlnLeuProLysThrThrGlnGlnSerGlnSerSerAlaInAsnAspMetProGln	2440
QY	7321	TGCGTGAATATGTCCTCTGAAAGATTTAACAAGTGGAGTGCCTGCAACTTTCAAAAAAG	7380
Db	2441	SerLeuAspIlyrValProGlnGlnPheAsnLysCysGlnCysProGlnLeuSerLysLys	2460
QY	7381	GGATCTATGATTCATACAAAAAAATTTACTGAAACCTAAATATCCTATGAAATTTGTGTAG	7440
Db	2461	GlySerMetIleHisThrLysLysIleThrGlnProLysIleIleProMetAlaInCysValGln	2480
QY	7441	AAAGCAGATATTTATTTATCTAAGAAGACGAAATATATGATATTTACCTTGAAGGAA	7500

Db	2481	LysAlaIaIeTyrTyrLeuSerLysGluAlaGluAsnMetAspIleThrLeuLysGlu	2500
QY	7501	AAATTTATACCTTTGAGCTGTCACAAAGGAAAGAAAGTAAATATGTTGACTAATAT	7560
Db	2501	LysPheIleProIleGluSerThrTyrGluLysGluSerLysAsnSerTrpHisAsn	2520
QY	7561	AATCCTTGGCATCTTAAGAAACCTTATGCACCTGATTAATATATAGAGAGAAACCTT	7620
Db	2521	AsnProCysAspProLysLysProTyrAlaProAspLysTyrIleGlyArgAsnPro	2540
QY	7621	TGTGAAAAATAGAGAAAGAAATCGTTTTAAAGTGTATATATATGAAATGTTACAAAAT	7680
Db	2541	CysGluAsnArgIleGluLysAsnArgPheLysValAspTyrIleTrpLysCysTyrLysAsn	2560
QY	7681	TCAAACTTCTATACAGAGGAAAAAAGAGTATGTGTCCCTCAGAGAGAAACATATGTGC	7740
Db	2561	SerLysPheTyrIleGluLysLysAspValCysValProIleArgArgGluHisMetCys	2580
QY	7741	TTAAGAAATTTAGATGAATTTAAATTTGAAAGACTTAAGGATATGATATATCTCTAA	7800
Db	2581	LeuArgAsnLeuAspGluIleLysIleGluArgLeuLysAspSerAsnTyrLeuLeuLys	2600
QY	7801	ATGCTTCTGCAACTGACAGAAATGAAAGCATAGACATAATAAAAACTTCAACTCAGAC	7860
Db	2601	MetValArgArgThrAlaArgAsnGluGlyIleAspIleIleLysAsnPheAsnSerGlu	2620
QY	7861	AACGGTGGGCATGGAATCCAAATATGTGATACATATGAAATATATGTTCCGTGATCAGGT	7920
Db	2621	AsnGlyCysAlaMetLysAsnProIleCysAspThrMetLysTyrSerPheAlaAspLeuGly	2640
QY	7921	GACATAGTTAGAGAACAGATATGTTATACGAATTTGCTGTACTTACCTCCCTGTAAGATA	7980
Db	2641	AspIleValArgIleThrAspMetLeuArgIleGlyGlyTyrLeuProProValGluIle	2660
QY	7981	AAATTATATAGCTTTTTCATATCATATATATGAAATATGAGAAATATAAAATTAAGGTAGA	8040
Db	2661	LysLeuTyrLysValPheGluTyrIleTyrGlyLysTrpArgAsnLysAsnLysGlyArg	2680
QY	8041	AATAAATACAAACATGATGACAAAGGTTTCGTTTCGTTGGGATGGTATAGAAAAGAT	8100
Db	2681	AsnLysTyrAsnAspAlaGlnTrpPheArgSerAlaTrpTrpAspAlaAsnArgLysAsp	2700
QY	8101	ATTTGAAAGCAATGACGTGTCAAGACACCGAAGATGCAAAACGTTTTACAAAAGCAAGA	8160
Db	2701	IleTrpLysAlaMetThrCysLysAlaProGluAspAlaLysLeuPheArgLysGlyArg	2720
QY	8161	ATGATGATTTTGAACGCATACATTATATCAAGATATAGTGTGACATTAAAGACGATCA	8220
Db	2721	MetAspGlyPheGluArgIleThrLeuIleGluAspLysCysGlyHisLysAspAspPro	2740
QY	8221	CCGTGTGATGATATATATACCTCAACGGTTTCGATGGATGATCAATAGGTGTGAATATAT	8280
Db	2741	ProValAspAspTyrIleProGlnArgPheArgTrpPheThrIleTrpSerGluTyrTyr	2760
QY	8281	TGTAAAGCACTGATGAGACAATTTGAAAAATTTAAAAATCATGTGATCTGTAAACA	8340
Db	2761	CysLysAlaLeuMetGluGluLeuGluLysPheLysLysSerCysAspHisCysLysThr	2780
QY	8341	TCTGACAGATGCAAGAAATGATTATGATGAAATTAAGTGTGACAGTGTAAACGACATGT	8400
Db	2781	SerAspArgCysLysAsnAspTyrAspGluAsnLysCysGluIncLysLysThrArgCys	2800
QY	8401	CAGGATATATAAAATTTGTTCTTAAATGGAATTCCTATATGATATACAAATCAATAAA	8460
Db	2801	GlnGluTyrTyrAsnPheValLeuLysTrpLysSerLeuPheAspIleGlnSerAsnLys	2820
QY	8461	TACAAAGAAATTTGATGACAAACCAATATATACAAAATCTCTACTATGATCATGTTTCAA	8520
Db	2821	TyrLysGluLeuTyrGluGlnProIleTyrThrLysIleSerThrTyrAspHisValGln	2840
QY	8521	AATTTTGTACAAAAGTTGAAAACGTTTTAAAGATGTAATCTGTGTGAGACGTTTTCTGAA	8580

Db 2841 AspPheValGlnLysLeuYsrThrPheLysSerGluCysSerValGlnSerPheSerGlu 2860
 QY 8581 TATCTTCATGAACAAGTAAGTGTGTAATTAATTAAGAAATGATGCTTCC 8640
 Db 2861 TyLeuHISgluThrSerLysCysLeuasnTyLysPheasnGlnasnAspLysSer 2880
 QY 8641 AATATACGAACATATGCTTCGAAAGAACACCAAAAAGTTATAAGACCTGCAGTTGT 8700
 Db 2881 AsnIleArgThrTyAlaPheGlnGluThrProLysSerTyLysGlnAlaCysSerLys 2900
 QY 8701 ACACATCTCTAAGAAATCCATGATATGCTTACCGATCAAAACAAAGATGAT 8760
 Db 2901 ThrLeuProSerLysasnProLysasnProLysasnProLysasnProLysasnProLys 2920
 QY 8761 AAGGATTCGAACCTTTACCTTCTGCTGCAAGAAATGATTAATTAATTAATTAAT 8820
 Db 2921 LysGlnLeuGlnThrPheThrPheCysSerLysasnAspTyAspAsnAsnLeuasn 2940
 QY 8821 TCGAAGCGATACCTGTTCTTAATAGTTCAAGATTAACAAAGCTGATTCCTCA 8880
 Db 2941 TrpAsnAlaTyLeuValLeuasnSerSerAspAsnLysGlnValLeuIleProPro 2960
 QY 8881 AGAAGAACATTTATGTACAAGACCTATCACTGATATTAATTAAGAAAGTGATAA 8940
 Db 2961 ArgArgArgHisLeuCysThrArgProIleThrAlaTyAsnTyArgLysGlyAspLys 2980
 QY 8941 GAAATTTAAAAAAAACCTTCTACTCTCTGCTTCAAGCAAGCAATTTAGTCA 9000
 Db 2981 GluIleLeuLysLysLysLeuLeuThrSerAlaPheSerGlnGlnLeuGln 3000
 QY 9001 AATATTAATCGAAGAAAGTGTGCTTGAGCAATGAATATTAATTAATTAATTAAT 9060
 Db 3001 LysTyLysSerGlnGlnGlnLeuCysPheGlnAlaMetLysTySerTyAlaAspTy 3020
 QY 9061 TCCATTAATTAATTAAGCACTGATATGATGACACTTCATTAATTAATTAATTAAT 9120
 Db 3021 SerAspIleIleLysGlnThrAspMetMetAspThrSerLeuSerGlnLysIleLysLys 3040
 QY 9121 ATATTGAACATCAAAATGAAGCAACCGAAATCGTAAACATGATGGGAAATTAATA 9180
 Db 3041 IlePheGlnThrSerAsnGlnAlaThrGlnAsnArgLysThrTrpProLysasnAsnArg 3060
 QY 9181 CGTCAGATATGGCAAGCTATGTTATGCGATATAAAATGCTTCAATTAAGCAATTA 9240
 Db 3061 ArgGlnIleThrPheHisAlaMetLeuCysGlyTyLysIleAlaThrSerLysValThrLeu 3080
 QY 9241 GATGAAGAGTGTGTCATTAATTAAGCAAGCAAGCAATTAATTAATTAATTAAT 9300
 Db 3081 AspGlnGlyTrpCysGlnLeuProLysAspGlnGlnThrAsnGlnPheLeuArgTrpLeu 3100
 QY 9301 ATGATATGGCAAGCAAGCAATTAATTAAGCAAGCAATTAATTAATTAATTAATTA 9360
 Db 3101 IleGlnTrpAlaLysGlnIleCysLysGlnLysLysHisValSerAspSerLeuLysThr 3120
 QY 9361 AATATCCCTCGTCAAAACGAAGATTAATTTGAAGCTCAGAAATTTAAAGCAACCTGA 9420
 Db 3121 LysCysProArgSerAsnGlnLysAsnPheGlnAlaSerGlnLeuIleArgLysProLys 3140
 QY 9421 TGTGATGATTAATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9480
 Db 3141 CysGlnAsnAspIleArgLysTyLysSerLeuasnIleLeuIleLysAsnThrMetGln 3160
 QY 9481 AATCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9540
 Db 3161 AsnLeuasnIleLysTyLysGlnLeuLysAspGlnSerSerGlnLysAsnIleAspAsnLys 3180
 QY 9541 CCATCTGAAGAAATGTTCACTCATATTAATTAATTAATTAATTAATTAATTAATTAAT 9600
 Db 3181 ProSerGlnGlnValGlnSerTyLysSerLysAspSerGlnCysAlaLeuGln 3200
 QY 9601 TTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9660
 Db 3201 LeuasnAspIleasnGlnIleValThrGlyThrLysasnAsnGlnLysLeuPheLys 3220

QY 9661 GAAGTACAAAAAATTAATATCCGTTTATTAATTTGTTGAAGATGAACACACAAAAAT 9720
 Db 3221 GluValLeuLysLysLeuTyProGlyLeuTyPheValGlnAspLysThrHisLysasn 3240
 QY 9721 CATGTAAGTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 9780
 Db 3241 HisValLeuasnAspLysasnIleLysGlnGlnGlnThrValArgProLysAlaLeuTy 3260
 QY 9781 TTTCTTACCCCATGATGATTTCTTCTATCAAGACCTTATATTCACACATCGAGTA 9840
 Db 3261 PhePheThrProHisValAspSerPheTyGlnAlaProLeuPheSerThrHisArgVal 3280
 QY 9841 GCACAAATGATGATCCATAAATGATATTAATTAATTAATTAATTAATTAATTAATTA 9900
 Db 3281 AlaGlnTyAspProLysasnAspIleLeuLysSerSerIleSerValIleValSer 3300
 QY 9901 GCGTTAGTTTGTATACCGCTTCATTTCAATGAAGAAAAAATTCAAATCGTGGACTTG 9960
 Db 3301 AlaLeuGlyLeuIleAlaLeuHisPheMetLysLysLysPheLysSerSerValAspLeu 3320
 QY 9961 TTGCGTATCTGAATATCCCGAAGAGATGTAATGCTTACCTGATTCGAAATCAAAAT 10020
 Db 3321 LeuArgIleLeuasnIleProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 3340
 QY 10021 AGGTACATACCATTAATGAAGTGTCCATTAAGCAACAAACATATTAATTAATGAAGA 10080
 Db 3341 ArgTyLysProTyArgSerGlyProTyLysGlySerTyThrTyLysMetGlnGly 3360
 QY 10081 GATPACTAGTGAAGATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10140
 Db 3361 AspThrSerGlyAspGlnAspLysTyMetTrpAspLeuSerSerSerAlaThrSer 3380
 QY 10141 TCCGAAATGAGTATGAAGAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10200
 Db 3381 SerGlnSerGlnTyGlnGlnLeuAspIleasnAspIleTyValProGlySerProLys 3400
 QY 10201 TATAAACATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 10260
 Db 3401 TyLysThrLeuIleGlnValValLeuGlnProSerLysArgAspIleProSerAspAsp 3420
 QY 10261 ACACCAAGTATGATTAACACACGATGATGATGATGATGATGATGATGATGATGATG 10320
 Db 3421 ThrProSerAsnAspThrProArgThrAsnArgPheIleAspAspGlnTrpAsnGlnLeu 3440
 QY 10321 AACATGATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10380
 Db 3441 LysHisAspPheValSerGlnTyLeuProAsnThrGlnProAsnAsnAsnTyLysSer 3460
 QY 10381 GCACATTTCCAAATGAATACAGAACCTAATCTTATTAATTAATTAATTAATTAATTA 10440
 Db 3461 AlaAspIleProMetAsnThrGlnProAsnThrLeuTySerAspAsnProGlnGlnLys 3480
 QY 10441 CTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10500
 Db 3481 ProPheIleIleSerIleHisAspArgAspLeuTyThrGlyLysGlnIleSerTyAsn 3500
 QY 10501 ATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10560
 Db 3501 IleAsnMetSerThrAsnThrAsnAsnAspIleProMetAsnAlaArgAsnAspSerTy 3520
 QY 10561 AGAGGATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10620
 Db 3521 ArgGlyIleAspLeuIleasnAspSerLeuValValLeuasnLeuIleTyMetMet 3540
 QY 10621 AAGTAT 10626
 Db 3541 LysTy 3542
 RESULT 2
 ID 087326 PRELIMINARY; PRT; 3287 AA.
 AC 087326;

DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE Plasmid (Fragment)
 GN TM180VAR2.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM180;
 RX MEDLINE=1927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT Identification of a conserved Plasmodium falciparum var gene
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL; AJ420411; CAD20867.1; -
 FT NON_TER 3287 3287
 SO SEQUENCE 3287 AA; 383550 MW; 58F8F66FC244536 CRC64;

Alignment Scores:

Seq. No.:	0	Length:	3287
Score:	12954.50	Matches:	2440
Percent Similarity:	80.66%	Conservative:	267
Best Local Similarity:	72.71%	Mismatches:	524
Query Match:	66.59%	Indels:	125
DB:	5	Gaps:	43

US-10-087-013-1 (1-10628) x 08T326 (1-3287)

QY 34 ATGGGAATGACATCATCATATAGAGGAGATGCTAAAGCCCTATTATTAAGAAAGT 93
 Db 1 MetGlyAsnAlaIAsSerLeuGluGlyAspAlaLysSerProIleLeuLysGluSer 20
 QY 94 CACAAAAGTCAGAAATGTTTGAACGTTATGCCAAAATATTAAGACATCCATCAAA 153
 Db 21 HisLysSerLarGAsnValLeuGluArgTyrAlaLysAsnIleArgHisProSerLys 40
 QY 154 TATGCAAAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 213
 Db 41 TyrAlaLysGluHisValAspSerLeuLysGlyAspLeuThrLysAlaGluPheArgGly 60
 QY 214 GGGCTTCCTGCGCAGTAATTAAGCAATTAATTAATTAATTAATTAATTAATTAATTAAT 273
 Db 61 GlyProSerThrProValAsnLysHisAsnTyrTyrProLysProCysAsnLeuAsp 80
 QY 274 CATAGGAACATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 333
 Db 81 HisLysGluHisThrAsnLeuArgTyrAspAspValAsnLeuArgHisProCysHisGly 100
 QY 334 AGAACAACAAACGATTTGATGAAGATGAAGATGATGATGATGATGATGATGATGATGAT 393
 Db 101 ArgGluGlnAsnArgPheAspGluAspGluGluSerGluLysGluLysIleArgAsn 120
 QY 394 TATTAAGAAAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 453
 Db 121 TyrLysArgLysAsnAspAlaIleAlaCysAlaProCysArgAlaGlnHisMetCysAsp 140
 QY 454 AAAAAGCTGAGACCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 513
 Db 141 LysAsnLeuGluAlaLeuAsnTyrIleAsnThrGlnAsnIleHisAspLeuLeuGlyAsn 160
 QY 514 GTCTAGTTCACGAAATTCGAAAGGATGATGATGATGATGATGATGATGATGATGATGAT 573
 Db 161 ValLeuValThrAlaLysTyrGluGluGluSerIleValAsnAsnHisProHisLysGly 180
 QY 574 ACTTCAGACCTGCTACTGCTGCGAAGATTTTCAGATATAGGTGATTTGATGATGATGAT 633
 Db 181 ThrSerAspValCysThrAlaLeuAlaArgSerPheAlaAspIleGlyAspIleValArg 200
 QY 634 GGAATAGATATGTTTAAACCAATGTCATGACAAAGTACAAAGGAGGCTCCGAGAGGT 693
 Db 201 GlyLeuAspMetPheLysProAsnValHisAspLysValGluLysGlyLeuArgGluVal 220

QY 694 TTCAAGAAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
 Db 221 PheLysLysIleHisAspGluMetGluGluValLysAsnTyrTyrAsnProAspGly 240
 QY 754 TCTGGAATATTAATTAATTAAGAAAGCATGGGTGGAACCTGTAATTAAGAAATTAAGTGC 813
 Db 241 SerGlyAsnTyrTyrLysLeuArgGluAlaTyrTrpAspValAsnArgAsnLysValTyr 260
 QY 814 GAAGCTATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
 Db 261 GluAlaIleThrCysGluAlaLeuProLysSerAlaTyrPheLeuGlnSerGluAspAsn 280
 QY 874 ACACCATTAATTTCAAACTCTAAATGCCCAT---AAACAGAAAGGTTCTTACCAAT 930
 Db 281 LysGluLeuPheLeuTyrProLysCysGluHisAsnAsnLysAsnAspLeuProThrAsn 300
 QY 931 TTGATTAATGCTCCCTCAATTTTACGTTGCTGAGACAAAGGAGAAAGATTTTGGCCGA 990
 Db 301 LeuAspTyrValProGlnTyrLeuAlaGlyTrpPheAspIleTyrPheLysGluPheCysArg 320
 QY 991 AAAAGAAATATTAATTAAGAAAGTCAAGACCTCTGCTATGACAAAGACCTTA 1050
 Db 321 LysArgAsnIleLysLeuLysValLysAspSerCysCysAsnAspLysGluAlaGluLeu 340
 QY 1051 TATTTGATGATTAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1110
 Db 341 TyrCysCysHisAspGlyTyrAspCysThrThrIleThrLysLysGluSerLeuHis 360
 QY 1111 TTGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1170
 Db 361 LeuAspAsnLysCysThrAspCysLeuThrLysCysLysValIleGluValIlePheGly 380
 QY 1171 AATCAACAAAGACATTTAAACAAACAAACAAACAAACAAACAAACAAACAAACATCAT 1230
 Db 381 AsnGluGlnIleAlaPheLysGluLysGluLysTyrLysGluLysGluLysGluLysTyr 400
 QY 1231 TTTATGCAACGATTAACAAATTTGTCATTAATTAATTAATTAATTAATTAATTAATTAAT 1290
 Db 401 ValSerAsnAspAlaLysPheValSerAsnIleAsnSerGluTyrTyrLysGlnPheHis 420
 QY 1291 GAAAACTTAAGAAACCGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAATTAAGCAAT 1350
 Db 421 AspIleLeuArgAspLysAsnHisAsnAsnLeuAspThrPheLeuAsnLeuLeuAsnGlu 440
 QY 1351 GGAAGATATTTAAAGGAGATTAACAGAGAAAGGATTTACTTTACTTAACAGTGT 1410
 Db 441 GlyLysTyrCysLysGluLysLeuLysGluGluSerAspIleThrPheThrAsnSerSer 460
 QY 1411 GATGCAAAAGGATTTTATTCGTTCAAGATTAATGCAAGTGTGTCCGACCTGCGGGGCT 1470
 Db 461 AspAspLysGlyIlePheTyrArgSerGlnTyrCysGlnValCysProAspCysGlyVal 480
 QY 1471 AAATGATGATGATTAATTAATTAACACACAAATTCAGATTAATGATGATGATGATTAATTAAT 1530
 Db 481 LysCysAspIleThrLysTyrThrHisLysSerAspAsnAspArgGluCysValAsnAsn 500
 QY 1531 GAAGACTATTAACCTCATGGGTGAGAACCTACTATTAATCACTGTCTTTATAGTGT 1590
 Db 501 GluAspTyrLysProProArgGlyValLysProThrAsnIleThrValLeuTyrSerGly 520
 QY 1591 AATGCAACAGTGTATTAACAAACAAATTTAGAAATTTTGTAAACGCTCACTAATTAAC 1650
 Db 521 AsnGluGlnIleLysAspIleThrGlnLysLeuGluAspPheCysLysSerSerThrAspTyr 540
 QY 1651 AAAGATTAATTAATTAACAAATGGAATGCTATTAATTAAGATGAAATTAATTAATTAATGAT 1710
 Db 541 LysAspLysAsnAsnGlnLysTyrPheLysCysTyrTyrLysAspGluAsnIleAsnArgCys 560
 QY 1711 AAATGGAACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1770
 Db 561 LysLeuGluGlnAsnThrGluIleAsnLysAspAsnProLysIleThrSerPheHisAsn 580

```
QY 1771 TTTTGAATATGAGGTACATATTTAAGGAGTACTATTAAGTGAATGACAACTT 1830
    |||
Db 581 PhehegluLeuTrpValThrTyrLeuLeuArgSprThrIleLysTrpAsnAspLysLeu 600
QY 1831 AAAACCTGTAATTAATACACACAGCATTTGATATGATATGTAACACAAATGCTTA 1890
    |||
Db 601 LysThrCysIleAsnAsnThrAlaThrHisCysIleAspLysCysLysAsnAsnCysLeu 620
QY 1891 TGTATTGACAGATGGCTTAACAAAAGAAAGAAATGGAATAGTATTAAGAAAGCTTC 1950
    |||
Db 621 CysPheAspArgTrpValLysGlnLysGlnGluGlnTrpSerIleLysLysLeuPhe 640
QY 1951 ACAAAGAAAAGAAATACGCAATGCTATTAATAGTATTAATTAATCTTTTGAAGT 2010
    |||
Db 641 ThrLysLysAsnAsnValProGlnProTyrTyrThrAsnIleAsnAsnLeuPheGlnGly 660
QY 2011 TATTTTAAAGTATGATTAAGTAACTTGACAAAGATGACAAAGAAAGAAAGTATG 2070
    |||
Db 661 TyrPhePheLysValMetAspLysLeuAspLysAsnGlnAlaLysTrpLysGlnLeuMet 680
QY 2071 GAAAAATATAAAGAAAAAAATGAGTTTCCAAATTTGAAAAATTAAGGAGTATTTA 2130
    |||
Db 681 GluAsnIleLysLysLysSerGlnPheSerAsnLeuGlnAsnAsnArgAspTyrLeu 700
QY 2131 GAGAAATGACATAGACTCTTGTAGACTTAAAGAAAGAAAGTCCCGATGATGTAAGAC 2190
    |||
Db 701 GluAsnAlaIleGluLeuLeuLeuAspHisLeuLysGlnThrAlaThrIleCysLysAsp 720
QY 2191 AATATATACAAAGCAATGATGAAACATCCATATATGACAAACAAACCCGTGTGTTAA 2250
    |||
Db 721 AsnAsnThrAsnGlnLysGlnLysGlnThrSerHisAsnAlaThrThrAsnProCysValLys 740
QY 2251 CCTCGGAGGAGGACGCAACCCCACTAAATATTAAGAAATAGCACAATATCTTTAAAG 2310
    |||
Db 741 ProAlaGlnGluGlnThrGlnProThrLysAsnIleLysGlnIleAlaGlnTyrPheLysArg 760
QY 2311 AGTGCATACAGAGAGACAGCAATATCGTGTCTTCAATTAATTAAGAAAGGACACAGCA 2370
    |||
Db 761 SerAlaLysGlnGlnLysAlaArgAsnArgLysLeuHisLysLysLeuGlnLysAlaHisGln 780
QY 2371 GGTATATATAAACGTGGGGGTAGAGAAAGCACTCAAGGACAATTTATGTAATATG 2430
    |||
Db 781 GlyIleTyrLysArgGlnLysAlaArgArgLysAspPheLysAspAsnLeuCysArgIleMet 800
QY 2431 ATAAACATCTTAATCGTATCTGTGTTTCAATGAGCACTGAGTGGGCAAGGACACA 2490
    |||
Db 801 IleLysHisSerAsnArgAsnLeuLysPheSerAsnGlnLysProCysAspLysGlnLysThr 820
QY 2491 GGTGATGATTAACAACAGATTTGCTAGAGACTGAATGGAATGGAATGCGGAGACAC 2550
    |||
Db 821 GlyAspGlnLysIleGlnThrArgPheValValGlnThrGlnLysValAlaAspProGlnHis 840
QY 2551 ATGCGTAAAGATCAGAGAGATGTTATGCTGCTCAACAGACATATATGATCATCC 2610
    |||
Db 841 MetArgLysAspHisGlnLysAlaValIleMetProProArgArgArgHisIleCysThrSer 860
QY 2611 AATTGGAACATTTCAACAGGATGATCACCACATTAATGGAATATGTTGATGATGATTA 2670
    |||
Db 861 AsnLeuGlnHisLeuGlnThrAspHisIleProLeuAsnGlnLysIleValAlaAspLeu 880
QY 2671 GTTAATTAATTCCTTTTGGGGGAGATGCTTCTTATCAGCAAAATATGAGCAACAAAGATA 2730
    |||
Db 881 ValAsnAsnSerPheLeuGlnLysValIleLeuSerAlaLysTyrGlnAlaLysLysIle 900
QY 2731 ATAGAGATGTAATTAAGAAAGAAATTAACCTTAAGGCCCCCAAGAAAGTATGATCACCACAAA 2790
    |||
Db 901 IleGlnMetLysLysGlnLysAsnAsnLeuLysGlnLysProLysGlnValThrAspProLys 920
QY 2791 CACCAAGACAACTATCTGTGAGCTATACGTTACGTTTGGAGATATAGGATATATAT 2850
    |||
Db 921 HisGlnAlaThrIleCysArgAlaIleArgGlySerPheAlaAspIleGlnLysAspIleIle 940
QY 2851 CGAGGAAGAGATCTCTGGGAAAGAAACGGTACATGTTAAAGCTCAAGACATTTGGAA 2910
    |||

Db 941 ArgGlnArgAspLeuTrpGlnLysArgAsnGlnLysAspMetValLysLeuGlnHisLeuGln 960
QY 2911 ACTGTTTGGTAATATATACATAGTCACTCAAGGCAAGAAAGAAATGATTAATATATGAT 2970
    |||
Db 961 ThrValPheGlnLysAsnIleTyrLysSerLeuLysGlnLysGlnLysAsnAspLysTrpAsnAsp 980
QY 2971 GATGCCCCCAATATTTAAATTTGAGGCAAAATTTGGTGGGAAGCTPATATAGCCCAAGTA 3030
    |||
Db 981 AspAlaProLysTyrLeuLysLeuArgGlnAspTrpTrpGlnAlaAsnArgAlaLysVal 1000
QY 3031 TGGGAGCCCATGAAATGTGATATTAATTTGAAGGATTAATGCGGACACCAACACACA 3090
    |||
Db 1001 TrpGlnAlaMetLysCysAspIleLysTyrLeuLysAspLysSerArgProIleSerThr 1020
QY 3091 CAAAGTATTTATGCGGATATATGATATCATAACCATTTGATGATTAATATCCACAAA 3150
    |||
Db 1021 GlnSerSerTyrCysGlnLysTyrSerAspHisThrProLeuAspAspTyrIleProGlnLys 1040
QY 3151 TTAAAGTATGATGACCGAATGGCGAGAAATGCTATCCAAAGTGCAGAAAAGGAGTATGAT 3210
    |||
Db 1041 LeuAlaGlyPheMetThrGlnTrpAlaGlnTrpTyrCysLysValGlnLysGlnLysTrpAsp 1060
QY 3211 AAGTTGAAGAGCAAGTGTAAAGAGTGTAAAGCATTAAGTAAATGCTCAAGGCTGTACGAAA 3270
    |||
Db 1061 LysLeuLysGlnLysCysLysGlnLysCysLysAspLysLysArgAsnGlnGlnLysCysThrLys 1080
QY 3271 GAGAGTGTATACAGTTGTGTACAAAGTGCACAGAGCTTGTAAATATATATATATATA 3330
    |||
Db 1081 GluSerGlyThrCysCysThrLysCysThrAspAlaCysAsnGlnLysTrpAsnAspIleIle 1100
QY 3331 GCATATGGAAGAAACAAATGGAATATATATATATATATATATATATATATATATATATATAT 3390
    |||
Db 1101 GlyLeuTrpLysGlnGlnTrpAsnIleIleSerAspLysTyrLysGlnLeuHisGlnGln 1120
QY 3391 GCACAAATGCTGTATGTAATATGATGTAATGTAAGCTTCCACTACTGCCCAAAATCATATA 3450
    |||
Db 1121 AlaGlnMetSerValSerAsnSerGlnLysIleAlaSerSerAlaIleLysAsnHisIle 1140
QY 3451 GACAGGAATGTTATGAAATTTTGTGCGAATTAATACCAACAAATGGTGGCAAAAGTAT 3510
    |||
Db 1141 AspArgAsnValIleGlnPheLeuSerIleLeuLysGlnGlnAspArgLysLysSerAsn 1160
QY 3511 AAAAGTGTACTAGTATGTAAGTGTGTCTCATTTGCTACTTAACACAGTATGAAATGTT 3570
    |||
Db 1161 LysSerGlyThrSerAspGlnSerThrValIleGlnHisAsnThrArgLysGlnAsnVal 1180
QY 3571 GGAGCATATCTCCATGATATACAGGAAATTTGATGATGTCATCACAATAATGAGTTTGT 3630
    |||
Db 1181 GlyAlaTyrLeuHisAspThrGlnLysAspAspAspCysGlnLeuGlnAsnGlnPheCys 1200
QY 3631 GATGAAAAAGTGAATGATGTAAGATTAACGAAATATATGCTTTTGAAGTAAACCAAGGAC 3690
    |||
Db 1201 AspGlnLysSerAspGlnLysAspLysGlnLysTyrAlaPheArgAspThrProAsnAsp 1220
QY 3691 CATGATGCTGCTGTGTTGTAATGATGTAAGTGAATGCAACCGCAAGGATGATTAATAACG 3750
    |||
Db 1221 HisAspValAlaCysAlaLysAsnGlnLysArgAspLys-----IleGlnProProLys 1237
QY 3751 AAAAAAAGCGGAGAAAGAAAGATACGGAATGTAAAAACAGTGAATGATTAATAAGAA 3810
    |||
Db 1238 GluLysLysLysGln-----AlaCysAlaIleValIleGlnLeuLeuLysGln 1253
QY 3811 AACGATGAAAGAAACAAAGTGAAGATGTCATCCCAAAAAAGAAATGATATGATATGCC 3870
    |||
Db 1254 AsnAsnGlnLysThrThrLysValGlnLysCysAsnValLysThrGlnGlnLys-----LysTrp 1272
QY 3871 GATTGGCAATGC-----GGAATATATATATATATATATATATATATATATATATATATAT 3918
    |||
Db 1273 GluTrpGlnLysGlnLysLysIleGlnAsnIleHis-----LysGlnAlaCys 1288
QY 3919 ATGCCCCCTAGAGACAAAGTATATGCTATATCTTGGCAATATATATATATATATATATAT 3978
    |||
```



```
Db 1289 MetProAlaArgPheLeuValHisPheLeuLysGlu-----Leu 1304
QY 3979 AAATTACATCAGAAAGTTAATAAAGAGCTTTCATCAATCTGCAGCAGAGAACA 4038
Db 1305 LysValGluThrValGluLysLeuArgGluAlaPheIleIncysSerAlaIaGluThr 1324
QY 4039 TTTCTTCATCGATTTATTAATAAGTAGAGTGAGCAAAATCTGCATTAAGAA 4098
Db 1325 PheLeuLeuTrpLysLysGluAspAsnGlnLysGluAspLeuGlnAsnGln 1344
QY 4099 TTAAGAAGCAAAATCTCCCGCATTTTGGATCCATGTTCTACATTTGGAGAT 4158
Db 1345 LeuGluSerLysIleIleProAspPheLysArgGlnMetPheTrpPheGlyAsp 1364
QY 4159 TATAGACATTTTATTTGAGACAGATATATCAAAAGTCATGGTGAGGAAGTAAC 4218
Db 1365 TyrArgLysPheLysLeuGlyAspIleGlySerAspValGly-----AsnVal 1381
QY 4219 AAAGAGCAATAGATTTCTTTTCAAAAATGCGACCAAAAATCTCTAATGCAAAA 4278
Db 1382 AsnGlnLysIleAsnSerAlaLeuLysArgIleGlyGlnSerAspGluAsp----Lys 1399
QY 4279 CGCCAGACATGTGAGCAGACAATAGTATGATGAGAGTATGATGACACTA 4338
Db 1400 ArgLysAsnTrpIlePheAsnGlyLysGluAspValIleTrpLysGlyMetValCysGlyLeu 1419
QY 4339 -----GTAAATAATGGGCAAAAAGATGAT 4365
Db 1420 SerHisAspValSerLysAsnHisLysGluIleValAlaArgLysLysIleMetGluAspPro 1439
QY 4366 TTTTACCAAAATCTACGTTTACACACACGTCATTAATAGTACAAA-----AGCACACT 4419
Db 1440 GlnAsnAsnLysLysArgLysAsnSerValLysTrpThrAspGluProSerGlyThrLys 1459
QY 4420 TTGGAGCAATTTGGCAACACACACCCAGTTTTCAGATGGCTAACCGAATGCTACAGAC 4479
Db 1460 LeuSerLysPheAlaLysValProGlnPheLeuArgTrpLeuThrGluIleTrpLysAsp 1479
QY 4480 TATTGCTATACAGACAAAAATATTGAGAGTGTGCAGAAAAATGTAATGTAATGAC 4539
Db 1480 TyrCysHisThrArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1499
QY 4540 CAATTGAAGTGTATACAGATGTAATGAAGAACGACACTACGTTAATATATATAA 4599
Db 1500 GlnLeuLysCysAspTrpGlnLysAsnLysLysCysGluAspLysLysLysLysLysLys 1519
QY 4600 AAA-----AAAAAGAGTGTATCCACAGATTAATATATACAGATGAGACGACAAA 4656
Db 1520 LysGluLysLysGluTrpIleLeuGlnAspLysTrpLysAspGluLysAspLysLys 1539
QY 4657 AGATTGATAGACACACACATTTGTTAATGTTACAGACTTACTGAGACGAATGACACA 4716
Db 1540 ArgPheAsnArgGlnHisIleGlyValIleMetValLysAspTrpThrGlyThrAsnAlaThr 1559
QY 4717 GATTACTTGAACGAATTTACTGCTAGTTGTGCTATAGCCTGGAAGGCTCTGTG 4776
Db 1560 AspTrpLeuLysArgLysPheThrAlaSerCysGlnAspLysProGlyGlySerSerVal 1579
QY 4777 GTACAAAGAAATATACATTTGTAAGAAACAGGCTTACTATGATGCGCAAAACATGTT 4836
Db 1580 ValGlnArgAsnIleGlnMetLeuGlnLysGlnAlaLysTrpLysAspLysHisCys 1599
QY 4837 GGGTGCAAAATTTATTTGAATGACAGACAAATATATACATTTGAGTAAAGATAG 4896
Db 1600 GlnCysThrLysPheIleGlnAsnAspAspLysTrpThrAsnIleSerSerLysAspAsn 1619
QY 4897 TGCAAAGATTTAGTAAGGAGGCAACACAGCTCTATTAGTGGCAAAACAAAGTCTT 4956
Db 1620 CysLysGlyLeuValLysGluAlaAsnThrGlyAlaLysTrpGlnAsnLysGlyPro 1639
QY 4957 AATACTACATTAATCTGAAAGAAATGACAGATGATGCTTTTCTTCTCGCGACTA 5016
Db 1640 AsnAsnTrpAsnAsnLeuLysGluLeuThrGlnLysPheLeuPheProSerArgArgLeu 1659
QY 5017 CGTATATGTTTTCATGCTATGATGGCAATTAACAGATCCAGAGTTAAAGATGAAT 5076
Db 1660 ArgIleCysPheHisAlaLeuAspGlyAsnTrpThrAspProGluValLysAspGluAsn 1679
QY 5077 GGGTGGCAAAAGATTTGATGAGAGTGCGCAACGACGATACATTTGGTCAATAC 5136
Db 1680 GlnLeuArgLysArgLeuMetGluValAlaAlaThrGlnLysLysAsnLeuGlnTrp 1699
QY 5137 TACAAAGAAAAAAGAAAAAGAAAAATAAAAACGTCGATGCGCAAAATTTCTAT 5196
Db 1700 TyrLysGluLysLysGluLysGluLysLysLysLysLysLysLysLysLysLysLys 1719
QY 5197 GAGTCCCGCTTGTAGTGTATGAAATATAGTTTATGATTAAGATATATATCTA 5256
Db 1720 AspValGlnProCysSerAlaMetLysTrpSerPheTrpAspIleArgAspIleIle 1739
QY 5257 GGTATTTATTAATTTGGAAAGTGAAGAAAAACCAACGACGAAAAATTTGAAAGAAAT 5316
Db 1740 GlnTrpAspAsnLeuGlnLysAspLysGlnAspIleGluLysAsnMetLysLysIlePhe 1759
QY 5317 AACAAATGGAACATCAGTTGGCAAGAAAGATGATGATGATGATGATGATGATGAT 5376
Db 1760 SerLysAsnArgThrSerValValLysGlnSerAspSerSerThrGlyAsnProGlySer 1779
QY 5377 ACTGCGCAAAATTTTCTGGAACGAAATTAAGGAATGTGTGGAACGCAATGATATGC 5436
Db 1780 SerValAlaTrpThrGluPheTrpAsnLysAsnLysGluValIleTrpAsnAlaMetLysCys 1799
QY 5437 GGGTACCAAACTGTAGGAGTATGGAATATGCGAAATGCTGCAACAGTATGAAGAT 5496
Db 1800 GlnTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1811
QY 5497 CTAAATAATGTGTTTCTGATCTTCAAGATGATATATCTATGAGGAAAAATCGCAT 5556
Db 1812 LysLysLysGluGluValPro-----AspAspLysLysLysLysLysLysLysLys 1828
QY 5557 GAAAGTCTGCTGATACGTTCTTGCATGCTTTCGCAATGCGGTGGAAGATTTTGGCAA 5616
Db 1829 GlnGlyThrAlaLysGlnPheLeuArgTrpPheThrGluTrpGluAsnAspPheCysLys 1848
QY 5617 CATTAAGAAAAAGAAATGGAATTTGTTAGGGCGCTGATGATGATGATGATGATGAT 5676
Db 1849 HisLysLysLysGluLeuGlnLysLeuValGlyAlaCysAsnAspTrpThrCysAlaAsp 1868
QY 5677 AATGAGATTAAGAAAGAAATGATACAGATGCTGATACAAATTAATAAATTTATTAAT 5736
Db 1869 AsnLysAspLysArgLysLysCysThrAspAlaCysThrGlnTrpLysPheIleSer 1888
QY 5737 GAGTGGAAACCACTATGAAAAACAATCAAAATATGCTGAGAAATTAAGACAAATA 5796
Db 1889 GlnTrpLysProGlnTrpGluLysGlnIleLysLysTrpGlyGluAsnAsnAspGluIle 1908
QY 5797 TATTCGAGCATCTGCGCAAAAGATGCAAGAGCTGCGCAATTTTAAGACAAACAA 5856
Db 1909 TyrSerLysAsnProValAlaLysGlyAlaLysAspAlaArgAspTrpLeuAspLysGln 1928
QY 5857 TTAATAAATAATTTGTGAAATTAAGAGAGATGTAATATATAGTATGAAAGATGTG 5916
Db 1929 LeuGlnLysPheCysGlnAsnLysSerLysAspCysGluTrpArgCysMetLysAspVal 1948
QY 5917 TCCACACAGCGATTAATGATGTAATAGTCAAAATATGCCCGCATGATTAAGACATGAA 5976
Db 1949 SerThrGlnSerLeuThrAspGlyAsnSerGlnAsnMetProAlaSerLeuAspAspGlu 1968
QY 5977 CCAAAAGAGTGAAGGAAGTGAATTTGCAATGCGACAGAGTCCACACAGCTGTACA 6036
Db 1969 ProLysGluValGlnGlyLysCysAsnCysGlnValProThrGlyArgProAlaArg 1988
QY 6037 AGGGAACACCGTCCACAGGGGTATCACTGATATCAAAAGACGCGATGAAAGAAAGAA 6096
Db 1989 ArgGluThrProSerProGlyAlaProLeuThrAlaLysAlaThrTrpSerGluLysGlu 2008
```


Db 2702 TyrAsnAspHisProProValasPaspTyrIleProGlnArgLeuArgTyrMetLysGlu 2721
 QY 8266 TGGTCTGAATATATATGTAAGACACGATGAGAAATGGAATAATTAATAATCATGT 8325
 Db 2722 TrpGlyIleValCysLysIleLeuAsnGlnLysValAsnHisValArg 3092
 QY 8326 GATCAGCTGTAACATCTGAC---AGATGCAAGATGATATGAAATTAAGTGGAA 8382
 Db 2742 GlnLysCysLysLeuAsnAspLysCysSerAspAsnAspArgLysLysCysArg 2761
 QY 8383 CAGTGTAAACGAGATGTCAGAAATATATAATTTGTCCTTAATGGAATCTCTATG 8442
 Db 2762 SerCysLysGlnLysCysLysGlnTyrThrLysLeuLeuHisAsnLeuLysSerGlnPhe 2781
 QY 8443 GATATACATCAATTAATATACAAAGAAATTTGATGACACCAATATATACAAATCTCT 8502
 Db 2782 AsnIleGlnLysLysLysTyrAsnGlnLeu-----TyrThrLysIleGln 2796
 QY 8503 ACTTAT-----GATCATGTTCAAAATTTTGTACAAAG 8535
 Db 2797 AsnAsnArgArgGlyPheIleAsnAspAsnAspLysAsnValIleGlnPhePheGlnLys 2816
 QY 8536 TTGAAAACCTTTAAAGTATGTCGTGTGAGAGCTTTCTGATATCTTCATGAAACA 8595
 Db 2817 ValLysMetLeu---AsnAsnCysAsnValGlyThrProAspLysTyrLeuAspLysAla 2835
 QY 8596 AGTAAGCTGTTGAATATTAATTAATTAATGAATGATGCTTCCTCAATATACGAACATAT 8655
 Db 2836 IleHisCysIleHisTyrAsnPheThrGlnAsnGlyLysSer-----LysProTyr 2853
 QY 8656 GCTTGGCAAAACCAAAAGTATTAAGAACTGTCAGCTGTACACATCTCTACAG 8715
 Db 2854 AlaPheAsnAsnHisProGlnLysTyrLysSerHisCysSerCysThrIle---ThrHis 2872
 QY 8716 AATCATTTGATATATGTCCTACCGATCAACAAAGATGATGATGATGATGATGATGATGAT 8775
 Db 2873 HisProLeuAspLysCysProAsnTyrLysThrAsnAlaTyrCysLysThrIleArgHis 2892
 QY 8776 TTTACCTTGCCTCGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8835
 Db 2893 IleAsnProCysIleThrLysAsnPheAspAsnAsnLeuIleThrThrGlnPheVal 2912
 QY 8836 GTTCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8895
 Db 2913 ValAspAsnLysLysAspLysAsnLysGlyValLeuValProProAlaGlnArgHisLeu 2932
 QY 8896 TGTACAGACCTATCAGCTGAT 8955
 Db 2933 CysThrArgLysLeuThrGlnLysAsnArgTyrArgIleAsnGlnLysAspAsnLeuLysGln 2952
 QY 8956 AATCTTCTACTCTGCTTCTCAGTCAAGACAAATTTGAGGTCAAAAATATATATATATATATAT 9015
 Db 2953 AsnLeuIleAspSerAlaPheSerHisGlyMetLeuLeuGlnLysThrPheAsnAspTyr 2972
 QY 9016 GAAGAGTTGCTTGGGCGAATGAT 9075
 Db 2973 ThrAsnGlnLysMetGlnSerMetLysTyrSerPheAlaAspTyrAlaAspIleLys 2992
 QY 9076 GGAACGATATGATGACACTTCA-----TTATCTGCAAAAATATATATATATATATATATAT 9123
 Db 2993 GlyThrAspMetIleGlyLysSerAsnIleAspAspPheAsnLysAspLeuLysGlnMet 3012
 QY 9124 TTGGAACATCAAT-----GAGCAACCGAAATCGTAAACATGCTGG 9168
 Db 3013 PheProGlnLysAsnSerGlnAsnIleGlyLysThrThrIleSerArgGlnGlnTrpTyr 3032
 QY 9169 GAAATATATAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9225
 Db 3033 GlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 3052
 QY 9226 TCAAAAGTAT 9285
 Db 3053 AsnAsnGlnGlyLysMetLysLysAsnTyrProCysAsnValProThrGlnLysArgLysThrAspGln 3072

QY 9286 TTTCTTCTGTTGTTAATGTAATGGCGCAAGACATGTAAGAAAAGAAACATGTAGT 9345
 Db 3073 PheLeuArgTyrLeuIleGlnTyrPalaMetGlnLacCysLysValLysAsnHisValArg 3092
 QY 9346 GATTCATTAATAACAAATGCTCGTTCGTAACAGACATATATTTGAAGGTCGAGATTA 9405
 Db 3093 AspSerLeuLysThrLysCysArgCysSerLysGlnLysAsnPheGlnLysLeuLeu 3112
 QY 9406 TTAAGACACCTGATGTCAGAAATGATATATATATATATATATATATATATATATATATATAT 9465
 Db 3113 LeuArgGlnProGlnCysGlnAsnAspIleArgLysTyrIleSerLeuAsnIleLeuIle 3132
 QY 9466 AAAATATCAATGCAAAATCTTAAT 9525
 Db 3133 LysAsnSerMetGlnAsnLeuAsnIleLysTyrLysGlnLeuLysAspGlnProSerGly 3152
 QY 9526 AATATAGCAATTAACATGATGAGAAATGTCAGATCATATATATATATATATATATATATATAT 9585
 Db 3153 AsnIleAspAsnLysProSerGlnGlnAsnValGlnInseryTyrIleLysSerLysTyrSer 3172
 QY 9586 CAATGCGCTTGGAGTAAATGAT 9645
 Db 3173 GlucylAspLeuGlnLysLeuAsnAspIleAsnGlnLysLeuAspThrGlyThrLysAsnAsnGln 3192
 QY 9646 AATATGAATTCAAAGAAGTACTAAATAATATATATATATATATATATATATATATATATATATAT 9705
 Db 3193 AsnAsnGlnPheLys---ValLeuLysLysLeuTyrProGlyLeuTyrPheValGlnAsp 3211
 QY 9706 GAAACACACAAATAATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9765
 Db 3212 GlnThrHisLysAsnHisValLeuAsnAspLysAsnIleLysGlnGlnGlnGlnGlnGlnGlnGln 3231
 QY 9766 CCTAAGACACTCTATTTCTTTACACCCCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9825
 Db 3232 ProLysValLeuTyrPhePheThrProHisValAspPheSerTyrGlnArgLeuPheLeu 3251
 QY 9826 TCAACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9885
 Db 3252 ProThrTyrArgValGlnLysTyrAspProLysAsnAspIleLeuLysSerSerIleSer 3271
 QY 9886 GTTGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9933
 Db 3272 ValValIleValSerAlaLeuGlnLysLeuIleAlaLeuLeuPheMetLys 3287
 RESULT 3
 ID 08T325 PRELIMINARY; PRT; 1615 AA.
 AC 08T325;
 DT 01-JUN-2002 (TEMBLrel. 21, Created)
 DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
 DE 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
 GN TM284VAR3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM284;
 RX MEDLINE=21927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babbiker H.A., Raza A.;
 RT Identification of a conserved Plasmodium falciparum var gene
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL: AJ420412; CAD20688.1; -;
 FT NON_TER 1615
 SO SEQUENCE 1615 AA; 188639 MW; FCA896C00B86DEA6 CRC64;
 Alignment Scores:
 Pred. No.: 0 Length: 1615
 Score: 7367.00 Matches: 1358
 Percent Similarity: 87.22% Conservative: 69

Best Local Similarity: 83.01% Mismatches: 169
Query Match: 37.87% Indels: 40
DB: 5 Gaps: 12

US-10-087-013-1 (1-10628) x Q81325 (1-1615)

QY 34 ATGGGGAATGCAGCATCATTTAGAGGAGATGCTAAAGCCCTATTATTAAGAAAGT 93
DB 1 MetGlyAsnThrGluSerSerLeuGluGluAlaArgSerProSerIleIleGluSer 20
QY 94 CACAAAAGTCGACAGAAAGTTTGGACGTTATGCCAAAAATATAGACATCCATCAAAA 153
DB 21 GluAsnSerProArgAsnValLeuGluValArgValAlaLysAsnIleArgGlnAlaSerLys 40
QY 154 TATGCAAAAAGCATGTGGATCGTTGAAGGGGATTTGACAAAAGCAAAATTTGGTGGT 213
DB 41 AspGluAsnGlnHisValAspSerLeuLysGlyAspLeuThrLysAlaIleuPheArgGly 60
QY 214 GGTCTCTTACGCCAGTAATTAAGCATATTTATTTATTCATATCCATGTAATTTAGT 273
DB 61 GlyProSerThrProValAsnLysHisAsnLysLysLysLysLysLysLysLysLysLys 80
QY 274 CATAGGAACAATCACTAATTTACCGTATGATGATGATGATTTAGACATCCCTGGCATG 333
DB 81 HisLysGlnHisThrAsnLeuArgTyrAspAspValAsnLeuArgHisProCysHisGly 100
QY 334 AGAGAACAACAACCGATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 393
DB 101 ArgGluGlnAsnArgPheAspLysLysLysLysLysLysLysLysLysLysLysLys 120
QY 394 TATTAAGAACAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 453
DB 121 TyrLysArgGluAsnAspAlaIleAlaCysAlaProProArgArgHisMetCysAsp 140
QY 454 AAAAAGCTTGAAGCTCTAATGATGATGATGATGATGATGATGATGATGATGATGAT 513
DB 141 LysAsnLeuGlnAlaLeuAsnAspIleAsnThrGlnAsnIleHisAspLeuGlnLys 160
QY 514 GTACTAGTTACGACAAATACGAAGTGAATCAATTTGTAATATCATCCATCAAGAA 573
DB 161 ValLeuValThrAlaLysLysLysLysLysLysLysLysLysLysLysLysLys 180
QY 574 ACTTGAGAGCGCTTATGCTCTTGCACAGAAATTTGGCATATGATGATGATGATGAT 633
DB 181 ThrSerAspValAlaCysThrAlaLeuAlaArgSerPheAlaAspIleGlyAspIleLeuArg 200
QY 634 GGAATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 693
DB 201 GlyIleAspMetPheLysProAsnValHisAspLysValGluLysGlyLeuArgGluVal 220
QY 694 TTCAAGAAAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
DB 221 PheLysLysIleHisAspLysLysLysLysLysLysLysLysLysLysLysLysLys 240
QY 754 TGTGGAATTTATTAATAATTAAGAAGCATGATGATGATGATGATGATGATGATGAT 813
DB 241 SerGlyAsnTyrLysLysLeuArgGluAlaArgPheAsnValAsnArgAsnLysValLys 260
QY 814 GAAGCTATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 873
DB 261 GluAlaIleThrCysGlyAlaLeuProLysSerAlaTyrPheMetGlnSerGluAsnAsp 280
QY 874 ACACCATTAATTTCAAAATCTTAATGCGCCATTAACAAAGAAAGTTGCT--ACCAAT 930
DB 281 LysGlnLeuPheSerLysLysLysLysLysLysLysLysLysLysLysLysLysLys 300
QY 931 TTAGATTTATGCTCAATATTTAGCTTGTGCAAGATGGGAGAAAGTTTGGCGCA 990
DB 301 LeuAspTyrValProGlnLysLeuArgTyrPheAspLysLysLysLysLysLysLys 320
QY 991 AAAAGAAATATTAATTTGAAGAAAGCTCAAGACCTCTGCTGTAATGACAAAGACCTTA 1050
DB 321 LysArgAsnIleLysLeuLysLysValLysAspSerCysArgAsnTyrLysGlnArgGlu 340

QY 1051 TATGTAGCATTAATGACATGATTTGTACAGACACTAATTTGGAAGAAAGTATTTGCAT 1110
DB 341 TyrCysSerHisAsnGlyLysPheCysThrThrIleThrPargLysGlyIleLeuHis 360
QY 1111 TTGATTAATTAAGTACTGACTGTTCCGATAAATGCAAAAGTTTGAAGTTGGTTAGGG 1170
DB 361 LeuAspAsnLysCysThrAspCysSerThrLysCysLysValPheGluValLysLeuGly 380
QY 1171 AATCAACAAGAACGCTTTAAACAAAAGAAAATGTAAGAAATATCAATCATAT 1230
DB 381 AsnGlnGlnGluAlaPheLysLysGlnLysGlnLysLysLysLysLysLysLysLys 400
QY 1231 TTATGCAACGATTAACAAATTTGCTCAATATTAATTAATGATGATGATGATGATGAT 1290
DB 401 LeuSerAsnAspAlaLysPheValAsnAsnIleAsnSerGluTyrLysLysGlnPheHis 420
QY 1291 GAAAACTTAAGAAACGCAATATGCAACTAAATGCACTTTTAAATTTACTAAATGAA 1350
DB 421 AspGlnLeuArgAspLysAsnTyrLysAsnLeuAspThrPheLeuAsnLeuLeuAsnGln 440
QY 1351 GGAAGTATTTGAAGGAGGATTTACAGGAGAAAGATATTACTTTACTAAGCTGCT 1410
DB 441 GlyLysTyrCysLysGlnLysLeuLysGlyGluAsnAspIleAsnPheThrAsnSerSer 460
QY 1411 GATGACAAAGGATATTTATGCTTCAGAAATTTGCCAAGTGTGCCGACTGCCGGGCTC 1470
DB 461 AspAspLysGlyThrPheTyrArgSerGlnTyrCysGlnValCysProAspLysGlyVal 480
QY 1471 AATGTGATGATTAATTAATCAACACAAATATGATGATGATGATGATGATGATGAT 1530
DB 481 LysArgAspGlyThrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 500
QY 1531 GAAGCTATAAAGCTCCATGAGGCTGTAAGCTACTAATATGATGATGATGATGATGAT 1590
DB 501 GluAspTyrLysProProIlePheLysLysLysLysLysLysLysLysLysLysLys 520
QY 1591 AATGACAAAGGTGATTAATCAACAAAATTTAGAAAATTTTGTACAGCTCAACTAATAC 1650
DB 521 AsnAspGlnGlyAspIleThrGlnLysLeuGlnAspPheCysLysSerSerThrAsnTyr 540
QY 1651 AAAAGTAAATAAATCAAAAATGGGAATGCTATTGTAAGAGAAATTAATTAATGATGT 1710
DB 541 LysAspLysAsnAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 560
QY 1711 AAACGTGAACAATAATCAAGAAATCAATATGATGATGATGATGATGATGATGATGAT 1770
DB 561 LysLeuGlnGlnAsnThrGluLysAsnAsnAspAsnProLysIleThrSerPheHisAsn 580
QY 1771 TTTTGTGAATTAATGCTTACATATTTATTAAGGATACTATTAAAGTGAATGACAACTT 1830
DB 581 PhePheGluLeuThrValThrLysLeuArgAspThrIleLysTyrAsnAspLysLeu 600
QY 1831 AAAAGCTGATTAATAATCAACCAACGCTGATGATGATGATGATGATGATGATGATGAT 1890
DB 601 LysThrCysIleAsnAsnThrThrThrHisCysIleAspGluCysAsnArgAsnCysLeu 620
QY 1891 TGTTTGACAGATGGGTTAAACAAAAGAAAGAAATGGAATAGTAATAAGAACTGTTCC 1950
DB 621 CysPheAspArgTyrValLysGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 640
QY 1951 ACAAAAAAAATAATACAGCAATCGTATTATAGTAATTAATTAATTTTGAAGT 2010
DB 641 ThrLysLysAsnAsnValProGlnProLysTyrThrAsnIleAsnAsnLeuPheGlnGly 660
QY 2011 TATTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2070
DB 661 TyrPhePheLysValMetLysPheLysLeuAspLysAsnGlnAlaLysTyrLysGlnLeuMet 680
QY 2071 GAAATATTAATAAAGAAAAAAGATTTTCCATTTGGAATTAATTAATGAGACATTTTA 2130
DB 681 GluAsnIleLysLysLysAsnGlnLysPheSerAsnLeuLysAsnAsnArgAspLysLeu 700

Accession	Gene	Protein	Length	Score	Alignment
Db 1400	ALG1	ALG1	1419	1400	ALG1
QY 4339	ALG1	ALG1	1419	4339	ALG1
Db 1420	ALG1	ALG1	1419	1420	ALG1
QY 4366	ALG1	ALG1	1419	4366	ALG1
Db 1440	ALG1	ALG1	1419	1440	ALG1
QY 4420	ALG1	ALG1	1419	4420	ALG1
Db 1460	ALG1	ALG1	1419	1460	ALG1
QY 4480	ALG1	ALG1	1419	4480	ALG1
Db 1480	ALG1	ALG1	1419	1480	ALG1
QY 4540	ALG1	ALG1	1419	4540	ALG1
Db 1500	ALG1	ALG1	1419	1500	ALG1
QY 4597	ALG1	ALG1	1419	4597	ALG1
Db 1520	ALG1	ALG1	1419	1520	ALG1
QY 4657	ALG1	ALG1	1419	4657	ALG1
Db 1540	ALG1	ALG1	1419	1540	ALG1
QY 4717	ALG1	ALG1	1419	4717	ALG1
Db 1560	ALG1	ALG1	1419	1560	ALG1
QY 4777	ALG1	ALG1	1419	4777	ALG1
Db 1580	ALG1	ALG1	1419	1580	ALG1
QY 4837	ALG1	ALG1	1419	4837	ALG1
Db 1600	ALG1	ALG1	1419	1600	ALG1
RESULT 4					
ID 015870	PRELIMINARY	PRT	2706 AA		
AC 015870					
DT 01-JAN-1998	(TREMBLrel. 05, Created)				
RT 01-JAN-1998	(TREMBLrel. 05, Last sequence update)				
RL 01-DEC-2001	(TREMBLrel. 19, Last annotation update)				
DR EMBL	(Fragment).				
DR EMBL	R29R+VAR1.				
DR EMBL	Plasmodium falciparum.				
DR EMBL	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
DR EMBL	NCBI Taxid:56833;				
DR EMBL	(11)				
DR EMBL	SEQUENCE FROM N.A.				
DR EMBL	STRAIN-IT 4/25/5;				
DR EMBL	MEDLINE=97373957; PubMed=9230440;				
DR EMBL	Rowe J.A., Kouids J.M., Newbold C.I., Miller L.H.;				
DR EMBL	"Plasmodium falciparum rosetting is mediated by PfEMP1 and requires				
DR EMBL	complement receptor 1.";				
DR EMBL	Nature 388:292-295(1997).				
DR EMBL	EMBL; Y13402; CAA73831.1; -				
DR EMBL	EMBL; Y13403; CAA73831.1; JOINED.				
DR EMBL	InterPro; IPR002086; Aldenhyd_dehydrt.				
DR EMBL	InterPro; IPR004258; PfEMP.				
DR EMBL	Pfam; PF03011; PfEMP; 2.				
DR EMBL	PROSITE; PS00070; ALDENHYD_DEHYDR_CYS; UNKNOWN_1.				
DR EMBL	NON TER				
DR EMBL	FT				
DR EMBL	SEQUENCE				
DR EMBL	2706 AA; 308162 MW; 1C3D55AD5317D68 CRC64;				

QY 3157 TGCATACCGAATGGCGAATGCTACTGCAGAGTGCAGAAAAAGAGTATGAATTG 3216
Db 722 -----Cysasp11eLeuProLysSerValAspAsp 732
QY 3217 AAGGAAAGTGTAGAGAGTGTAGATAGATAGATAGATAGATAGATAGATAGATAG 3276
Db 733 GUGLUAASPAPSPUGLUAVAL ----- 739
QY 3277 GGTACAGTTGTACGAAGTGCACAGAGCTGTGAATGAATATATATATATATATAT 3336
Db 739 ----- 739
QY 3337 TGGAAAGAAATGAT 3396
Db 740 -----AspGluGluLysGlu ----- 744
QY 3397 ATGTCTGTAGTATAGTGTATGAGTGTCCAGTACAGTCCAAAATCATATAGCAG 3456
Db 745 -----GluSerSerGluThrThrLysArgAsnIle ----- 754
QY 3457 AATGTATATGAATTTTGTGCAATTTATACAAATAATGTCGCAAAAGTATATAAGT 3516
Db 755 -----SerGluLysGly 758
QY 3517 GGTACTAGTATGAAAGTCTGTCTCATGTCTACTACACACGATGAAAAGTTGGAGA 3576
Db 759 G1YThrLysSerAlaSerCysValLysGlyAlaCysAlaIleValLysGlyVal ---Leu 777
QY 3577 TATCTCCATGATACAGAAATTTGATGTGTCAGTCAAAATGATTTGTGTATGA 3636
Db 778 G1NGluLysSerAsnLysSerIleAspAsnCysAsnAlaLysAsnArg ----- 793
QY 3637 AAAAGTATGATGATAGATATACAAAAAATATGCTTTAGATTAACACAGACCATGAT 3696
Db 793 ----- 793
QY 3697 GGTGCGTGTGTGTAAGTGTGATCGAAGCAGACAGGTACAGTATAAAACGAAAAA 3756
Db 793 ----- 793
QY 3757 AAAGCGAAGAAAGATATCGAATGTAAACAGTGAATGATATCTTAAGAAAAAGAT 3816
Db 793 ----- 793
QY 3817 GCAAAAGAAACAGTACAGATTTGATCCAAAAAAGATAGTAAATGCCGATTCG 3876
Db 794 -----LysLysAsn -----GluTrp 798
QY 3877 CAATCGGAAATATATAATTTAGTGAA ---GACCCTCGTGTGTATGCCCCCTTGAGAA 3933
Db 799 G1NCysAspLysAsnThrPheValAspGluAsnGluGlyValCysMetProProArgArg 818
QY 3934 CAAAAGTATGCGTACATTTCTGGCAATGATATGAAATTAATAAAATTCACATCA 3993
Db 819 LysSerIleCysIleHisAsnLeuThrLeuGluGluGluThrLys -----AsnLys 835
QY 3994 GTTAATTTAAAGAAAGCTTTTCACTCAATTCGACAGACAGAAACATCTTCATGCTAT 4053
Db 836 TyrGluLeuArgGluAlaPheIleLysCysAlaAlaLysGluThrAsnLeuLeuTrpAsp 855
QY 4054 TATTTAAAGTAAAGATGTTGAGAAATGAACTCGAATTAAGATTAAGAAAGCAAA 4113
Db 856 LysTrpLysAsnAspLysAsnGluAlaGluGluLeu -----LeuLysLysGlyLys 872
QY 4114 ATTCCTCCCGATTTTGAATTCATGTTCTACATTTGAGATTTAGATTTTGA 4173
Db 873 IleProGluAspPheMetArgIleMetPheThrThrPheGluAspPheArgAspPheCys 892
QY 4174 TTTGGAACAGTATATCAAAAGCTATGTCGAGGAACTAAACATTAAGACCAATATAT 4233
Db 893 LeuGluAsnAspMetGlyLys -----AspValAspLysValLysLysAsnIleAsn 909

QY 4234 TCTCTTTCAAAAATGTCGACCAAAATCTCCTAATGAAAAACACGCCAAGATGTGG 4293
Db 910 LysValAlaPheAsnAsnSerSerLysArgLysLysIleAspProLysAsnTrp 929
QY 4294 ACAGAACATATGATGATATGGAAGCTATGCTATGTCACATAGTAAAAATGGGCA 4353
Db 930 AsnGluAsnGlyProGluIleThrPheAsnGlyMetLeuCysAlaLeuIleHisAlaAspThr 949
QY 4354 AAA -----AAGATGATTTTACCGAAAACTACGGTTTACAAACAGTCGAA 4398
Db 950 LysAspSerIleLysAsnLysAsp -----AsnTrpLysTrpGluLysValThr 965
QY 4399 TTTAGTACAAA -----AGCACACTTTGAGGAATTTGCCAAGACCC 4443
Db 966 IleLeuAlaLysArgAspLysSerAsnGlyMetThrLeuSerGluPheAlaLysLysPro 985
QY 4444 CAGTTTTCATGCTGCTTAAACGATGATGACGACACTATGCTATACACAAAAATAT 4503
Db 986 LysPheLeuArgTrpPheValGluTrpLysAspArgLysGluLysGluLysTrp 1005
QY 4504 TTGAAGGATGTCAGGAAAAATGTAAGTCAATGAC -----CAATTGACGTGATACA 4557
Db 1006 LeuThrGluValAlaSerThrCysLysSerIleAspGlyGluLysLeuLysCysAspArg 1025
QY 4558 GAATGTAATTAAGAAATGCGAGGACTACGTTAATATATGAAAAAATAAA ---GAGTGG 4614
Db 1026 GlyCysAsnAsnLysCysAspLysTrpLysLysTrpMetArgLysLysLysLysLys 1045
QY 4615 ATTCACAAAGATATATTTACAGATGACCGCAAAATAAAAGTTGATAGACACAC 4674
Db 1046 AsnLeuGluAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1065
QY 4675 ATGTGTATGATGTTACAGACTATCTGTAACGATGTAACGATGTAACGATGTAACGAT 4734
Db 1066 IleGlyIleIleValLysAspTrpValLeuAlaAsnAlaLysGluThrLysLysLysLys 1085
QY 4735 TTTACTGCTAGTGTGTGATTAAGCTGGAAGTGC -----TCTGTGTA 4779
Db 1086 PheThrAlaSerCysValThrSerSerGlyLysAlaGluAsnSerAlaThrGluGluVal 1105
QY 4780 CAAAGAAATATATCTGTTGAATAACAGCGCTTCTATGATGCCGACAAACATTTGGG 4839
Db 1106 LysLysAsnIleGluLeuLeuSerGluGluGluGluGluGluGluGluGluGluGluGluGlu 1125
QY 4840 TGCACAAAATTTATGAAATGACGACAAATATATCTAATTTGCACTTAAGATTAAGTGC 4899
Db 1126 CysThrLysPheIle ---HisAspLysLysLysLysLysLysLysLysLysLysLysLysLys 1144
QY 4900 AAAGATTAATTAAGGAGGACAAACACAGTCTATTAAGTGCACAAACMAAGTCTTAAT 4959
Db 1145 CysGlyLeuAsnSerAspAlaLysLysAsnLysIleLysTrpArgAsnSerAspGluLys 1164
QY 4960 AACTCAATTAATCTGAAGAA ---TTGACTGAAGTGTGCTTTTCTCTTCGTCGA 5013
Db 1165 AspTrpAlaPheLeuLysLysArgAsnLeuSerGlyAspValPhePheProSerArgArg 1184
QY 5014 CTACGTATATGTTTTCATGATGATGCAATGCAATTAACAGATTCACAGAAATTAAGATGA 5073
Db 1185 LeuArgIleCysPheHisAlaLeuAspLysAsnTrpThrAspProGluValLysAspLys 1204
QY 5074 AATGGTTCGCAAAAAGTTATGGAAGTGGCGCACGGAAGGTATATTTGGTCAA 5133
Db 1205 AsnGlyLeuArgLysArgLeuMetGluValAlaAlaThrGluGlyTrpAsnLeuSerGln 1224
QY 5134 TACTACAAAGAAAAAAGAAAAAGAAAAATTAACAGTGGGACAAATATATCT 5193
Db 1225 TyrTrpLysGluLysLysGluLysGluThrGluAlaThrGluGluAlaHisLysTrpSer 1244
QY 5194 TATGAGTCCGCGCTTGTAGTCTATGAAATATAGTTTATGATTTAAGAGATATAT 5253
Db 1245 TyrGluValGlnProCysSerAlaMetLysTrpSerPheTrpAspLeuArgAspIleIle 1264
QY 5254 CTAGTATGATTAATTTGGAAGATGAAAAACAAAGACCGAGAAAAATTTGAAGAAAAATA 5313

[illegible]

QY 7012 ---AGTAAT-----GACGGTAAGACACAGAAATGT 7038
Db 1971 CysSerAsnIleAspValaenLysAspLysIlePheHalaAspLeuGluGlyProLysCys 1990
QY 7039 GCAGAGCGCTGTGTATACATTCATAATTTATTAAGAGAGGAGAAACTGAAATATGAAGA 7098
Db 1991 AlAlaYsProCysSerSerIYrAryGlyStrIleLysGlyLysLysThrGlnTyGluLys 2010
QY 7099 CAAGAGAGAAAGTTCAGAAAAGATTAAGATGGCAAAAAGTATAG-----GATTAT 7149
Db 2011 GlnGluLysAlaTyGluGlnGlnLysGluGlnCysLysLysGluThrAsnAsnHisTyr 2030
QY 7150 CCTTCTACCTAAGAGACATAGAGAGGACACATGTGCTCATGATATTTAAACATGAAGA 7209
Db 2031 AsnGlyPheArgValLysLeuGlyThrCysThrThrAlaGlyAspPheLeu----- 2047
QY 7210 TTAAGAAATTAATGTCATTAAGATTTGTTCTGTATGCAAAAACCTTTCACAACCTA 7269
Db 2048 -----GlnThrLeuLysAsnGlyProCys----- 2055
QY 7270 CCAAGAAACACACACACATCCATCCGATGCTAATGATATGCGACAGAAATGCTGAT 7329
Db 2056 -----LysSerGluAsnGlyLysAspHisGlu----- 2065
QY 7330 TATGTCCTGAAGATTTAAACAGTGTGAGTCTCCGACCTTCAAAAAAGGATCTATG 7389
Db 2066 -----AspGluLysAsnPheSerGln---ProAspValThrPheArgProAla--- 2080
QY 7390 ATTCAATACAAAAAATTAACCTTAATAATCTATGATGATGATGATGATGATGATGATGAT 7449
Db 2081 -----ThrAsnGlyThrProCysProLysPheLysValAsnGly----- 2093
QY 7450 TATATTTATCTAAAGACAGAAAAATTAATATGATATTAACCTTGAGAGAAAAATTTATA 7509
Db 2093 ----- 2093
QY 7510 CCTATGAGCTCTACAGAGAAAAAGAAATTAAGTGTGACATAATAATATCTGTGC 7569
Db 2094 -----LysAsnGlyAsnGlyAlaAspThrAsnGlyLysCys 2106
QY 7570 GATCCTACAGAAACCTTATGACCTGCTAATAATATATAGAGAGAAACCCCTGTGAAAT 7629
Db 2107 AsnGlyLysThrPro----- 2111
QY 7630 AGAGAGAAATCGTTTAAAGTATATGATGATGATGATGATGATGATGATGATGATGAT 7689
Db 2111 ----- 2111
QY 7690 TATCAGAGAGAAAAAGAGTATGTACTCCAGAGAGAGACATATGTCTTAAGAAAT 7749
Db 2111 ----- 2111
QY 7750 TTAGATGAATTAATTAAGAACTTAAGATAGTATATATCTCTTAATAATGTTGCT 7809
Db 2112 IleAspAlaGlnAsnPheGluGlnMetGlyGlnThrAlaLysGluPheValMetLeuVal 2131
QY 7810 CGAAGTGCAGAAATGAAGATAGACATTAATAAAAACTTACATCAGAGAGCGGTC 7869
Db 2132 SerAspLysSerThrAsnGlyPheGluVal-----AsnAspLeuAsnGluLys 2147
QY 7870 GCAATGAATCCAAATGTGATACATATGAATAATAGTTTCGTGATCGGTGATGATGAT 7929
Db 2148 Ala----- 2148
QY 7930 AGAGAGACAGATATGTATGATATGATGATGATGATGATGATGATGATGATGATGAT 7989
Db 2149 ---GlyAlaAsp----- 2151
QY 7990 AAGGTTTGAATCATATATGAGAAATGAGAGAAATTAATAAAGTATGAATTAATATAC 8049
Db 2152 ---IlePheGlnGlyLe----- 2156

QY 8050 AACGATGACAAAGCTTTCCTGCTGTGGATGCTAATAGAAAGATATTTGAAA 8109
Db 2157 -----LysGluAsnIleTrpSer 2162
QY 8110 GCATATC-----ACGTGCAAGACACAGAA-----GATGCAAAA 8142
Db 2163 CysArgAsnValCysGlyLeuAspValCysLys---ProGluLysValAsnAspGlnLys 2181
QY 8143 CTTTAAAGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8202
Db 2182 ValAsnGlyLys---GluAsnAspGly----- 2189
QY 8203 GGACATAGAGAGATCCACCTGTGATGATATATATCTCAACGGTTTGATGATGATGAT 8262
Db 2190 -----ThrTyrllelleGlnIleArgAlaLeuLeuArg 2200
QY 8263 GAATGCTCGAATATATTTGTAAGACAGTGAAGAGAAATGGAATAAATAAATCA 8322
Db 2201 ArgTrpValGluAsnPhe-----LeuGluAspTyrLysLysIleLysHisLys 2216
QY 8323 TGTGATCACTGTAAACATCTGACAGATGACAGATGATGATGATGATGATGATGATGATGAT 8382
Db 2217 IleSerHisCysThrAsnSerThrGluGluLys-----LysSerThrCysAsp 2232
QY 8383 CAGGTAAACGAGATGTCAGAAATATTAATAAATTTGTTCTTAATGAAATCTCTATTC 8442
Db 2233 ---CysGlyLysLysCysLys-----CysValGlyGlnTrpIleLysLeuLys 2247
QY 8443 GATATACATCAAAATTAATACAAAGATTTGTATGAAACACATATATACAAAAATCTCT 8502
Db 2248 LysGluGluTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2258
QY 8503 ACTTATGATCATGTTCAAAATTTGTACAAAAGTTGAAAACCTTTAAAGTGAATGTTCT 8562
Db 2259 -----ValGlyGluAsnLysSerGlu----- 2265
QY 8563 GTTGAGAGCTTTTCGAATATCTTATGAAACAGTAAGTGTGAAATTAATTAAT 8622
Db 2266 ---AspAspPheSerAspAsnLeuAsnSerPheLeuGlnThrIleThrGlnIlePro 2284
QY 8623 GAAATGATGTTCTTCAATATACAAATATATGCTTGAAGAGACCAAAAAAGTAT 8682
Db 2285 ValAlaAspValGlnGlyAsnValIleLysLeuSerAsnPheAspThrPro----- 2301
QY 8683 AAAGAGCTGTCAGTTGTACACTTACCTTGAAGATTCATGATTAATGTCTACCGAT 8742
Db 2302 -----CysGlyCysSerAlaAspAlaAsnSer----- 2310
QY 8743 CAAACCAAGATGATGTAAGGAATTAACAACCTTTACTCTGCTGAGAGATGATAT 8802
Db 2311 GlnLysLysAspGly-----AsnGluAsnAspAla 2330
QY 8803 GATATAATCTTGATTAATTTGAGAGCATACCTGTTCTTAATGATGATGATGATGATGAT 8862
Db 2321 IleAspCysMetIleAsn-----ArgLeuGlnGlnLysAlaLys 2333
QY 8863 GGTGATGATGTTCTCCAGAGAGAGACATTTATGATACAGACCTATCAGCATATTAAT 8922
Db 2334 AsnGlyGluThrGlnProSerGlySerLysGlnCysThrThrProThrThrLeu--- 2352
QY 8923 TATGAAAAAGGTGATTAAGAAATTTTAAAAAAAACCTTCTACTTCTGCTTCAAGTCA 8982
Db 2352 ----- 2352
QY 8983 GGACAAATGTTAGTCAAAAAATATATAATCGAGAGAGAGTGTGCTTGAAGCAATGAAA 9042
Db 2353 -----GluAspAspGluThrPheAsp----- 2359
QY 9043 TATAGTTATGAGATTTATTCGATATATTAAGAGAACTGATATGATGACACCTTCATTA 9102
Db 2360 -----AspAspIle----- 2362
QY 9103 TCTGAAAAAATTAATAAATTTTGAAGACATCAAAATGAGACACGAAAAATCGTAAACA 9162


```

Db 2363 -----GluThrAspAsn----- 2366
QY 9163 TGGTGGGAAATATAGACGTGACATATGCGACCGCTATGTATGATATATAATTCCT 9222
Db 2367 -----ProValAlaHisProGlnIleCysGluGlyValIleGln 2379
QY 9223 ACTTCAAAAGTAACTATAGATGAAGCATGGTGTCAATTCACCAAGATGAAGAACTAAT 9282
Db 2380 ThrGlnThrGluThrValIleGluGluGlyValIleProValGluGluGlyVal 2399
QY 9283 CAGTTTCTGGTGTGAATGAATGGGCAAGACATGTAAAGAAAGAAACATGTA 9342
Db 2400 Asp-----GluLysLysLysPro 2405
QY 9343 AGTGATTCATTAAACMAATGCTCTGTCAACGAGATATATTTGAAGCGTCAGAA 9402
Db 2406 GluAspPro----- 2408
QY 9403 TTATTAAGACACCTGATGTCAAGATGATTAAGAAATATATTAAGCTTGAATATATG 9462
Db 2408 ----- 2408
QY 9463 ATAAAAATATCAATGCAAAATCTAATATATAATGAATTAAGCAATTAAGATCAATCTCA 9522
Db 2409 -----AlaGluGluAspGly 2413
QY 9523 GGTAAATATAGACATATA-----CATCTGAGAAATAATGTCTCATATATATA 9570
Db 2414 GlyAlaIleGlyProSerGlyProAlaGlyProGlnIleProSerProIleHis 2431
QY 9571 AAATCAAAAGATTCCTCAATGCCCTTGAGCTTAATGATTAATGAATAGTATACAGA 9630
Db 2431 ----- 2431
QY 9631 ACAAAATATATGAATATGAATTCAGAAAGTACTAATAAAATATATATCTGCTTAA 9690
Db 2431 ----- 2431
QY 9691 TATTTGTGAGATGAACACACMAAATCATGTACTAGTGAATATATAAGAAAGAA 9750
Db 2432 -----AspAspThrGluGluAsnHisVal 2439
QY 9751 GAGCAAAAGTGTCTTAAGACACTATTTCTTACACCCCATGTAGATCTTTTGTAT 9810
Db 2440 -----ThrProGluGluAspProPro 2447
QY 9811 CAAGCACCCTTATTCACACACATCGAGTAGCACAA----- 9846
Db 2448 ProAlaPro-----AspThrArgProValProSerProLeuProProProAlaAspGln 2465
QY 9847 ---TATGATCTTAATAATGATATATATAAGTATCTCTGTGTATTTGATTCGCG 9903
Db 2466 ProAspAspPro-----ThrIleGluGlnThrThrIleProHeGlyValAlaLeuAla 2483
QY 9904 TTAGCTTTGATAGCGCTTCATTCATGACAGAAATAATTCAAATGCTGTGAGACTTTGTTG 9963
Db 2484 LeuGlySerIleAlaPhePhePheMetLysLysThrLysHisProValAspLeuPhe 2503
QY 9964 CGTATTCGATATCCCGCAAGAGAGATGATGAGATGCTGTGGAATCCAAATAAGG 10023
Db 2504 SerValIleAsnThrProLysGlyAspLysAspIleProIleLeuLysSerLysAsnArg 2533
QY 10024 TACATACCATATAGAAAGTGTCCATATATAAGCAAAACATATATATATGGAAGAGAT 10083
Db 2524 TyrIleProLysArgSerGlySerLysGlyLysThrLysIleLysMetGluGlyAsp 2543
QY 10084 ACTGTGAGAGATGAAGTAAATATATATGCGACTTATCTCTGATATATATCTATCC 10143
Db 2544 Ser-----AspGluAspLysLysLysAlaPheIleSerAspThrIleHisSer 2561
QY 10144 GAAAGTAGATGAGAAAGATGATATATATATATATATGACAGGATGCTCAATAT 10203

```

```

Db 2562 GluSerIuThrGluGluLeuAspIleAsnAspIleThrValProGlySerProLysTyr 2581
QY 10204 AAACATTGATAGAGTAGTACTAGAACCATCAAAAG-----GATATACCA---AGT 10254
Db 2582 LysThrIleuIleGluValIleGluProSerLysSerAsnGlyAsnIleProHisSer 2601
QY 10255 GATGATACCAAGTAATGAT---ACACACGTCAGCAATGATTTATGATGATGATG 10311
Db 2602 AlIGlyLysProLeuAspAspMetValProThrThrAsnThrPheThrAspGluGluTyr 2621
QY 10312 AATGACATGAACATGATTTGTATCTCATATATATATACCAATACAGAACCAATATAT--- 10368
Db 2622 AsnIleuLysHisAspPheIleSerGlnIleGlnSerArgLeuProMetAspVal 2641
QY 10369 ---AATTACAA---AGTCAGATATTCATGATGATACAGAACCTAATATATATCT 10422
Db 2642 ProGlnIleAspValSerThrGluLeuProMetAsnIleValGlyAsnValLeu---Asp 2660
QY 10423 GATATCTCGAAGAAACCTTTTATATATATATATATATGATGATGATGATGATG 10482
Db 2661 AspLysIleAsnGlyLysProPheIleThrSerIleHisAspArgAsnLeuLysAsnGly 2680
QY 10483 AAAGAAATGATATATATATATATATATATATATATATATATATATATATATATAT 10542
Db 2681 GluIleIleSerLysHisAsnIleAsnMet-----ValAsnAsnAspIleProMetSer 2697
QY 10543 GCTAGAAATGATTTCTTATAGAGTATA 10569
Db 2698 AspLysAsnGlyThrLysThrGlyLe 2706

RESULT 5
Q26032
ID Q26032 PRELIMINARY; PRT; 3006 AA.
AC Q26032;
DE 01-NOV-1996 (TEMBUREL. 01, Created)
DE 01-NOV-1996 (TEMBUREL. 01, Last sequence update)
DE 01-DEC-2001 (TEMBUREL. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes.";
RL Cell 82:89-100(1995).
DR EMBL; LA0609; AAA75397.1;
DR InterPro; IPR004258; PfEMP.
DR Pfam; PF03011; PfEMP; 2.
SQ SEQUENCE 3006 AA; 343769 MW; 8FD5FF475F943C74 CRC64;

Alignment Scores:
Pred. No.: 8.63e-207 Length: 3006
Score: 4260.00 Matches: 1143
Percent Similarity: 43.84% Conservative: 483
Best Local Similarity: 30.82% Mismatches: 1005
Query Match: 21.90% Indels: 1079
Gaps: 134

US-10-087-013-1 (1-10628) x Q26032 (1-3006)
QY 91 AGTACAAAGTGCAGCAATATGTTTGGACGTTATGCCAAATAATATAGACAT---CCA 147
Db 8 ThrAsnLysThrAlaLysGluValIleAspIleLysGluValGlnGluLysAla 27
QY 148 TCAAAATATGCAAAAGACATGTGATTCGTTGAAGGGATTTGACAAAGCAGATTT 207

```

Db 28 ThrGluAspAlaLeuThrTyrArgAsnAspLeuGlnGlyAsnLeuSerGlnAlaLysPhe 47
QY 208 CGTGGTGGCTCTTACCGCAGTAATAGCATATTTATTCATCCATCATGTAAT 267
Db 48 HisGly-----ValProIleAspValLysAsn-----ProCysAsp 59
QY 268 TTAGATCATAGGAACATACATTAATTTACGTATGATGTAATTTGAGACATCTTGC 327
Db 60 LeuAsnTyrGlnIleHisThrAsnVal---ValLysGlyArgLysGlnLysProCys 78
QY 328 CATGTAGAGAAACAAACGATTTGATGAAGATGAAGATCGATGCGAAATATAATA 387
Db 79 ArgGlyArgGlnGlnLysArgPheSerAspValLeuSerGlnGlnAlaLysAsnThr 98
QY 388 CGTAAATTAATAAGAAAAAATGATGCTATAGCTGTGCGCCACCTGAAGACGATATG 447
Db 99 IleLysAspSerValThrAsnSerValGlyAlaCysAlaProPheArgGlyLeuHisVal 118
QY 448 TGTGATAAAACTGGAAAGCTTAATGATTAATATCCAAATATTCATGATTTATG 507
Db 119 CysAspTyrAsnLeuGlnLeuIleLysThrAspGlnAsnThrSerThrHisAspIleLeu 138
QY 508 GGAATGTACTAGTTACAGCAAAATACAGAGTGAATCATATGTTATTAATCATCCAT 567
Db 139 ValAspValLeuLeuThrAlaLysHisGlnGlnLysSerLeuValLysHisTyrLysGln 158
QY 568 -----AAGAACTTCAGACGCTTACTGCTCTTCACAGCAAGTTTGCAGAT 615
Db 159 TyrIleLysLysAsnArgAsnPheAsnIleCysThrValLeuAlaArgSerPheAlaAsp 178
QY 616 ATAGTGATATTTGTAGAGAGATAGATATGTT-----AAGCAAAAT 657
Db 179 IleGlyAspIleValAlaArgLysAspLeuTyrLeuGlnTyrAspAspGlnGlnLysGln 198
QY 658 GTCCATGACAAAGTACAAAGCGGCTCCGAGAGCTTTCAGAAAAATACATGATGCAAT 717
Db 199 LysArgLysGlnLeuGlnLysAsnLeuLysPhePheGlnLysIleHisAspAspVal 218
QY 718 -----GAAGATGAAGTAATAATGATTCAT 744
Db 219 MetLysThrSerLysArgThrAsnGlnLysLysSerAlaGlnLysArgLysAsn 238
QY 745 CCGATGATCTGGAAATTTATTAATTAAGAAAGCATGTGTGAATGATAGAAAT 804
Db 239 ---AspAlaThrGlnLysAsnTyrLysLysLeuArgGlnAspTyrPheAsnAlaAsnArgAsp 257
QY 805 AAGATGTGGAAGCTATACATGATGATGATATTAATCTGATATTTATGCAATCA 864
Db 258 GlnValAlaTyrLysAlaIleThrCysAspAlaAlaAspAsnAspGlnTyrPheGlnLysSer 277
QY 865 GAAAGTAATACACCATTTTCAATCCTAATGCGCCATTAACAAAGAAAGTCTCT 924
Db 278 SerAspGlyLeuTyrValPheSerAsnGlnLysGlnArgAsnGlnLysValPro 297
QY 925 ACCATTTGATTAATGCTCCCAATATTTAGCTGTGCGCAAGTGGGAGAGAGATT 984
Db 298 ThrAsnLeuAspTyrValProGlnHisLeuArgTyrPheAspGlnTyrAlaGlnLysPhe 317
QY 985 TGCAGAAAAAGAAATTAATTAATGAAAAAGCTCAAGGACTCTCTGCGTAATGCAAAAGA 1044
Db 318 CysArgLysArgAsnIleThrLeuLysSerAlaLysGlnLysCysGlyGlyGlnLysAsn 1044
QY 1045 CCGTATATTTAGTCATTAATGACATGATGTCAGCAACTATTTGGAAGAAAGATTT 1104
Db 338 ThrLysTyrCysSerLeuAsnGlnLysAspCysThrLysValPheGlnLysLysAspSer 357
QY 1105 TTGCATTTGGAATTAATAGTACTGACGCTGCGCAATTAATGCAAGTTTGAAGTTTG 1164
Db 358 CysSerSerAspGlyAsnCysThrAlaCysSerAsnIleCysIleAlaTyrAspAlaTyr 377
QY 1165 TTAGGGAATCAACAAGAGCATTTTAAAAAAGAAAAATATGAAAAAGAAATCA 1224
Db 378 LeuArgAsnGlnArgAsnGlnPheGlnLysGlnLysIleLysTyrThrLysGlnIleGln 397
QY 1225 TCATATTATCGAAGCATTAACAATTTGCAATTAATTAATAGATATTATAACA 1284
Db 398 LysTyrLysSerSerSerAspLysSerAsnSerAsnIleSerAsnLysTyrTrpAsnGln 417
QY 1285 TTTTATGAAAAAATTTAAGAAAGCAATATGCAATTAATGACATTTTAAATTTACTA 1344
Db 418 PheTyrGlnLysAsnPheGlnLysLysGlnTyrGlnThrLeuGlnAsnPheLeuLysLeu 437
QY 1345 AATGAAGAAAGATTTGTAAGAGAGATTAACAGCAAGAAAGATTAATCTTTACTAAC 1404
Db 438 AsnLysGlyMetLysCysGlnGlnLysIleGlnGlnGlnValIleAspPhe----- 455
QY 1405 AGTGTGATGACAAAGGATTTTATGCTGAGAAATATGCAATATTTGCCAAGTGTCCGACATG 1464
Db 456 AsnLysAspGlnLysPheValPheHisArgSerGlnTyrCysGlnProCysProAspCys 475
QY 1465 GGGGTCAATGTATGATTAATTAATTAATGACACAAATGATGATTAATGCGTAACGT--- 1521
Db 476 ValValGlnCysLysGlnLysCysThrGlnAspLysLysAsnAspLysCysArgSer 495
QY 1522 -----GTAAATTAAGACGTATAAACCTCATGGGCTGTGAAGCT 1563
Db 496 LysIleIleLysLysIleLeuGlnSerGln-----Pro 507
QY 1564 ACTAATATCACTGCTCTTATAGTATGAAGCAAGGATATTAACAACAAATTAAGA 1623
Db 508 ThrGlnIleHisValLeuAsnSerAspAspLysGlnLysAspIleThrLysLysLeuGln 527
QY 1624 AATTTTGTACAGCTCAATTAATTAACAAGATTAATAAATAATGCAAAATGGAATGCTAT 1683
Db 528 ValPheCysSerSerThrThrAsnTyrGlnLysArgAsnValGlnLysTyrLysCysTyr 547
QY 1684 TTAAGGATGAAATATTAATAGATGTAACCTGCAACAAATACGAAATATGAT 1743
Db 548 AsnLysAsnSerAspTyrAsnAsnCysGlnMetAsnIleSerSerTyrLysAspSerThr 567
QY 1744 AATCTTAAGATTAATTCATTTCAATTAATTTTGGTATACATATTTATTTAGG 1803
Db 568 AspAlaAsnValMetLeuSerValGlnCysPheHisSerThrAlaLysAsnLeuIle 587
QY 1804 GATTAATTAATGGAATGACAAACTTAACCTGTATTAATTAATTAACA---ACCAGCAT 1860
Db 588 AspThrIleLysTyrGlnHisGlnLeuLysAsnCysIleAsnAsnThrAsnValThrTyr 607
QY 1861 TGTATTGATGATGATTAACAGAAATGCTTATGTTTGCAGATGGGTAAACAAAGAA 1920
Db 608 CysGlnSerLysCysIleLysAsnCysGlnCysTyrGlnLysTyrLysArgLysGln 627
QY 1921 GAAGATGGAATAGTATTAAGAAACCTGTCAAAAAAAGAAATATATACGCAATCGAT 1980
Db 628 HisGlnTyrGlnLysValLysAsnValPheGlnLysAsnAsnArgMetSerTyrIleTyr 647
QY 1981 TATGTAATTAATTAATTAATTTTGAAGGTTATTTTAAAGTATGATTAAGTAACTTGAC 2040
Db 648 TyrAsnAsnLeuSerAlaGlyAlaPheAspSerPheLeuPheGlnValMetPheAlaLeuAsp 667
QY 2041 AAACATGAA---GCAAAATGAAAGAACTTATGCAAAATTAATAAAGAAAAAAGAG 2097
Db 668 GlnAspGlnLysGlnLysTyrAspGlnPheThrGlnLysPheLys---LysLysPheGln 666
QY 2098 TTTTCAATTTGAAATTAATAGGACTATTTAGCAATTAATGCAATCTTGTAGAT 2157
Db 687 ProSerLysThrAsnThrProThrGlyLysSerGlnAspAlaIleGlnPheLeuAsn 706
QY 2158 CACTTAAGAAAGCTGCCAGATATTAAGCAATTAATTAAGCAAGAGATGTGAACA 2217
Db 707 HisLeuLysAspAsnAlaLeuThrCysArgAspAsnAsnSerAsnLysSerCysAspVal 726
QY 2218 TCCCATTAATCAACAACAACCGGTGTAACTCTGTGAGGACCGCAACCATTA 2277
Db 727 SerLysLysValLysThrAsnProCysGlnLysAsnProSerAlaSerAsnLeuVal 746

Db 1425 LysTyrSerTyrLysLysLeuasnGluSerProLysGlySerAsnLysValGluAspHe 1444
QY 4432 GCCAACGACCCCGAGTTTTCATGATGGCTACACGAAAGGTGACGCGCTATGCTATAC 4441
Db 1445 AlAlLysLysProGlnPheLeuAArgTyrPheLleGluTyrGlyAspGluPheCysAlaGln 1464
QY 4492 CGACAAAAATTTTGAAAGGATGTCAGGAAAAATGT-----MAGTCMAATGAC--- 4539
Db 1465 ArgGluGluLysGluAlaLysValLysValSerCysSerAspAlaLysAspTyrAspGly 1484
QY 4540 -----CAATGAGTGTGATACAGACATGTAATATAAGAAATCCGAGGACTACGTTAAATAT 4593
Db 1485 CysLysAsnThrLysSerAsnAlaSerCysValSerAlaCysLysValTyrGluAspTyr 1504
QY 4594 ATGAAAAAAGAGAGAGTGCATCCACAGATTAATATTACAGATGAAAGCGACGACAA 4653
Db 1505 IleThrLysLysLysVal-----GluTyrThrLys-----GlnLys 1516
QY 4654 AAAAGATTGATGACACACATTTGTGTATGTTTACAGACTTACTGA----- 4704
Db 1517 GlyLysPheAspAlaGluLysIle-----ThrAspLysGluGlyTyrGluGly 1532
QY 4705 -----ACGATGCAAGATTTACTTGAACAGAAATTTACTGCTACTGTGTGAT 4755
Db 1533 PheSerThrLysAspAlaSerGluTyrLeuLysLys-----Cys----- 1546
QY 4756 AAGCCTGGAGGCTCTGTGTGACAAAGAAATATACATTTGTAGAAAAAGCGCTTAC 4815
Db 1546 ----- 1546
QY 4816 TATGATCCGACAAACATTTGGTGACAAAAATTTAATGAACGCAATATACT 4875
Db 1547 -----LeuAspAspThrCysAsnCysMetGlnLysAsnAsnThrGluTyrTyr 1564
QY 4876 AACATTTCCAGTAAGATAGTCAAAAGATAGTAAGAGGCAACACAGGCTGAT 4935
Db 1565 AsnThr----- 1566
QY 4936 AAGTGCAAAACAAAGGTCTCTATATACATACATTAAGTGAAGATTTGACTGAAGTGG 4995
Db 1567 -----ProAsnLysThrTyrThrAsn-----SerAsnLeuGlnLysArgCysGlu----- 1581
QY 4996 CTTTTCCTCTCGTGCAGCTACGTATATGTTTCATGATGGATGGCAATTATACAGAT 5055
Db 1582 CysGlnProProGln-----Glu 1587
QY 5056 CCAAGAGTTAAAGATGAAAAATGGTTGCCAAAAAGATTGATGAAAGTGGCGCAACGAA 5115
Db 1588 ProProProGlyProGluGlyAlaArg-----SerAspSer 1600
QY 5116 GGGTACAAATTTGGTCACTACACAAAGAAAAAGAAAAAGAAAAATAAACGCTG 5175
Db 1601 Gly-----ProArgAspThrThrProArgProAlaGlySer 1612
QY 5176 GATGCGCAAAATATCTTATGAGGTCCGCTGTAGTGTATGAATATATAGTTTAT 5235
Db 1613 AspAlaArgSerAsnThrValProSerProProArgProAlaGlyAspThrValHis 1632
QY 5236 GATTTAAGGATATATCTCTAGTATGATATTTTGAAGATGAAAAAGAAAGCGGAC 5295
Db 1633 GluValAlaGluVal-----GlnGluGluGluGluGluGluAspGlu 1645
QY 5296 GAAAAATTTGAGAAAAATATTTAACAAGAAATGACATCAGTTGGCAAGGAAGTATAGT 5355
Db 1645 ----- 1645
QY 5356 ACTACAGGAATCCCGGTAGTACTGCGGAAAAATTTTCTGAGCAAAATAGAGATGT 5415
Db 1645 ----- 1645
QY 5416 GTGTGAACGAATGATATGGGGTACAAACGCTGTAGGATGATGAATAGTGAAT 5475
Db 1646 -----AspGlyAspLeuProGlu 1651
QY 5476 AGTGCAGAAAGTATGATCTAAAAAATGTTGTTCTGCTACCTGATGATGATAT 5535
Db 1652 Asp-----GlnAspGluValGluValAlaGlyAlaGluGluGluGluGlu 1667
QY 5536 CCTATGGGAAAAATGCGGATGAGTACTGCTATCACTTTCTTGATGAGTTTGCCGA 5595
Db 1668 -----LeuAspValGlyValAla-----ArgIleLeuArgIleArgThrAsn 1681
QY 5596 TGG-----GGTGAAGATTTTGCACAAACATAAAGAAAGAAATGAGAAATGGAGGGCG 5652
Db 1682 SerProAspGluAspGluAspGluAlaSerGluGluGlu----- 1694
QY 5653 TGTAAATGATTAATCTGTGTGATATGAAGTAAAGAAAGAAATGACAGATCGGT 5712
Db 1695 -----AspAspAspAspAspAlaGlnAspThrGluValThr 1708
QY 5713 ACACATATATATATTTATGATGAGTGAACACACATATGAAAAACAAATCAAAAA 5772
Db 1709 GlyGln----- 1710
QY 5773 TATGTGACAAATTAAGACAAAAATATATTCGAGCATCTGTGCAAAAGATGACAGAG 5832
Db 1711 -----GlyGluGluGluThrAlaGluAspHisGlnAsp 1721
QY 5833 GGTGCGAATATTTAGACAAACATTAATAAATTTGTGAAATTAATGAGATTTGT 5892
Db 1722 ThrThrGluThrValAspGln-----GluLysAlaGluGluAspGlyGlyGly 1740
QY 5893 GAATATTAAGTATGAAAGATGTGCACACGCGATTACTGATGATGATATGATCAAT 5952
Db 1741 Glu-----ThrProGlnLysGluThr----- 1747
QY 5953 ATGCCCGCATCTTAGACATGACCAAAAGAAAGTTGAAGAAAGTGAATTTGCAAGTG 6012
Db 1748 -----GlnProLysValGluValAsnProCysAsp----- 1757
QY 6013 CCACGAGTCCACACGCTGACGAAGGAAACACCGTCCACACGGGTATCTGATATCA 6072
Db 1758 -----IleValLys 1760
QY 6073 AAAGGAGGCGATCGAAAAAGAAAGCAACAGCGCGCTACAAACACCGCAAAAA 6132
Db 1761 ThrLeuPheThrThrGluThrLeuLysGluLacysProThrLysTyrValAsnGly 1780
QY 6133 GTGAAATCTAACACGAATGGGACACAAACAGAACCCGACGACGACACAA 6192
Db 1781 ArgGlu----- 1782
QY 6193 ACACGAAAGCAACATCAACACAAACAAACACAGATCTGAGTGGCACAAATGTAAAG 6252
Db 1782 ----- 1782
QY 6253 GCAATCTCTTGAATTAACAGATACGAGGGGTGAATAGAGGTTGTAATCCAAAAAG 6312
Db 1782 ----- 1782
QY 6313 TATGACAAATCTCTAAATGCGGTGTATGCTAGTAACTAAGAAATGAAGAAATGCG 6372
Db 1783 -----LysPheProAsnThrLysCysIle-----SerSerLysSerAspAlaSerGlySer 1799
QY 6373 ATATGATGCTCTCTAGCAAAAAATATGATTAATAATATACATATTAATATAT 6432
Db 1800 IleCysIleProProArgArgLysLeuThrLysLysLysLysLysLysLysLysLys 1819
QY 6433 -----GAACTGAAAAATAGCGTACATATAT-----ATAAAGAGGCT 6471
Db 1820 ThrValSerSerAspGlyGluThrThrThrProIleThrHisAspAlaLeuArgGluAla 1839
QY 6472 TTTATTAATGTGACGACATAGAAACCTCAATTTTGTGGTAAATATAT-----ATAATGAA 6528
Db 1840 PheIleGlnThrAlaIleValGluThrPhePheLeuThrHisArgTyrLysLysLysLys 1859

Db 2348 LysLysSerProCysIleasnGlyCysGlnAsnLysCysasnCysValGluIleTyrPile 2367
QY 8434 TCTCTATTCGATATACATACAAATTAATACAAAGAAATGTATGACACCAATATATACA 8493
Db 2368 GlnLysLysSerGluIlePylGlyValArgGlnIleTyr----- 2381
QY 8494 AAAATCTCTATTCATGATCATCTGTTCAAAATTTGTACAAAAGTTGAAAACCTTTTAAAGT 8553
Db 2382 ---IleasnGlnIleTyr----- 2385
QY 8554 GAATGTTCTGTGAGAGCTTTTCTGAATATCTTCATGAACAAGTAAGTTTGAATTAT 8613
Db 2385 ----- 2385
QY 8614 AAATTTAAGAAAATGATGGTCTTCCATATACGAACATATGCTTCGAAGAACACA 8673
Db 2386 ---ArgAspLysAsnSerAsn-----GlnLlePheGlu-----Val 2396
QY 8674 AAAAGTTATGAAGACCTTGCACTGTACACTCTTCAAGAAATCCATTTGCAATATGCT 8733
Db 2397 LysSerPheLeuGlnIleThrLeuIleProGlnIlePro-----Val 2409
QY 8734 CTTACCGATCAAAACAA---GATGGATGAAGAAATTAACAACCTTTTACCTTGCTGCG 8790
Db 2410 ValThrAspLysGlyLysHisAspSerLeuThrGlnLeuLysLysLeuLysCysAsn 2429
QY 8791 AAGATGATTAATGATTAATATCTCTGTAATTTGGAACGACATCTGTTCTTAATAGTCA 8850
Db 2430 CysSerGlnLysSerGlnAsn----- 2436
QY 8851 GATGATTAACAAAGGTGATGATTCCTCCACAAGAACATTTATGTACAAGCCTATC 8910
Db 2437 SerAsnGlnLysAspValAl-----Leucys----- 2445
QY 8911 ACTGATATTAATTAAGAAAAGGTGATTAAGAAATTTTAAAAAAAACTTCTTACTTCT 8970
Db 2446 -----LeuLeuLysLys----- 2449
QY 8971 GCTTCAGTCAAGCAATTTAGTCAAAAATTAATTAATTCGAAAGAGTTGCTTT 9030
Db 2449 ----- 2449
QY 9031 GAGCAATGAATATAGTTATGACATATTCCTCATTAATTAAGAACTGATATGATC 9090
Db 2449 ----- 2449
QY 9091 GACACTTATTAATCTGAAAAAATTAATAATTTGAAAACATCAATGAAACCAACGAA 9150
Db 2450 ---LeuGlnAspLysAlaLysAsn----- 2456
QY 9151 AATGCTAAACATGCTGGGAAAATTAATAGCTCAGATATGCGACGCTATGTTATGTGA 9210
Db 2456 ----- 2456
QY 9211 TATAAAATTTGCTACTTCAAAAGTATACATTAAGATGAAGATGCTCAATTAACAAAGAT 9270
Db 2457 -----Lys 2459
QY 9271 GAGAATACTAATCACTTCTGTTGTTAATGTAATGGCAAGCAAGCATGTAAAGAA 9330
Db 2460 GlnAlaSerGlyLle----- 2464
QY 9331 AAGAAACATGTAAGTATTCATTAATAAAACAAATGCTCTGTTCAACGAAGATATTTT 9390
Db 2465 ---ProCysProGlnThrThrSerGlnAsnPro 2474
QY 9391 GAAGCGTCAGATTAATTAAGAACACCTGATGTGCAATGATATAGAAAATATATATTCGC 9450
Db 2475 AspAspGlnAspLleLeuLeuGlnIle----- 2483
QY 9451 TTGAATATATTAATTAATAATCAAGAAATCTAATATATATATATATATATATATATATA 9510
Db 2484 ---GlnAsnProValGlnLleProAsnIle----- 2492

QY 9511 GATCAATCTTCAGTATATATAGCAATTAACCATCTGAAAGAAAATGTTCACTATATATA 9570
Db 2493 -----CysProLysValGlnIleProGlnProValVal 2503
QY 9571 AAATCAAAAGATTCATACATCCCTTTGAGTTAAATGATATTAAGAAATAGTACAGCA 9630
Db 2504 ---GlnGlnLysCys-----AspLleAlaGlnLleProSerLys 2516
QY 9631 AAAAAAATATGAATAATTAATTAATTAAGAAAGTACTAAAAAATTAATATATCTGTTTA 9690
Db 2517 GluSerSerThrGlnLysAsnSerGlyLle----- 2526
QY 9691 TATTTGTTGAAGATGAACACAAACAAATCATGTACTAGATGAATATATAAGAAAGA 9750
Db 2527 -----GlySerAsnSerGlnGlnAsnProArgSerLysProGln 2539
QY 9751 GAGCAAAACGTTCTGCTTAAGCACTTAATTTCTTAACACCCATGATATCTTCTAT 9810
Db 2540 GluIleProProProProThrSerSerGlnThrAspThrPro----- 2553
QY 9811 CAAGCACCTTTTCTCAACACATGAGTACCAATATGAT---CTTAAAT----- 9861
Db 2554 ---ProProAlaProProThrThrIleGlnProSerGlnAlaAspGlnProThrAsnSerIle 2572
QY 9862 ---GATATATGAAGATGATATCTGTTGTTATTTGATATGCGGCTTGTGATAGCG 9918
Db 2573 SerAspLleLeuSerSerThrIleProPheGlyIleAlaIleAlaLeuThrSerIleVal 2592
QY 9919 CTTCAATTCATGAAGAAAATTAATCAATCGTCTGAGCTTTGCTATACATCAATATC 9978
Db 2593 PheLeuPheLeuLysLysLysThrLysSerSerValAspLeuLeuValLeuAsnIle 2612
QY 9979 CCGCAAGAGAGTATGAATGCTGCTGATGAATCCAAATTAAGTACATACATATGA 10038
Db 2613 ProLysGlyLysIleGlyLeuPheProThrLeuLysSerSerAsnArgLysIleProTyrAla 2632
QY 10039 AGTGTCCATATTAAGCAAAACATATATATATATGAAGAGATACATGAGATGAGAA 10098
Db 2633 SerAspArgLysLysGlyLysThrLysThrLysThrLysThrLysThrLysThrLysThrLys 2650
QY 10099 GATAATATATGCTGGATATCTTCTCTCATATATCTTATCTCCGAAAGTATGATGAA 10158
Db 2651 GlyHisTyrTyrGluAsp-----ThrThrAspValThrSerSerGlnSerGlnTyrGln 2668
QY 10159 GAATGATATTAATGATATATATATATATGATACAGTGTCTTAATTAATAAATGATGAA 10218
Db 2669 GluLeuAspLleAsnAspLleTyrValProLysSerProLysTyrLysThrLeuIleGln 2688
QY 10219 GTAGTACTAGAACCATCAAAAGGATATACCAAGTAT-----GATACACCAAGT--- 10269
Db 2689 ValValLeuGlnProSerGlyAsnAsnThrThrAlaSerGlyLysAsnThrProSerAsp 2708
QY 10270 -----AATGATACACACGCTACGATATATATATATATATATATATATATATAT 10311
Db 2709 ThrGlnAsnAspLleGlnAsnAspGlyLleProSerSerLysIleThrAspAsnGlnTyr 2728
QY 10312 AATGAACATGAACATGATTTGTTATCTCAATATATTAACAAATAGCAACAAAT----- 10365
Db 2729 AsnThrLeuLysAspGlnPheIleSerGlnIleTyrLeuGlnSerGlnGlnProAsnAspVal 2748
QY 10366 ---AATATTAACAAAGTGCAGATATTCACATGAATACAGAACCAATATATATATCT 10422
Db 2749 ProAsnAspTyrThrSerGlyAsnSerSerThrAsnThrAsnIleThrThrSerArg 2768
QY 10423 GATAATCTGAAGAAAACCTTTATATATATATATATATATATATATATATATATATATAT 10482
Db 2769 HisAsnValGlnGlnLysProPheIleMetSerLleHisAspArgAsnLeuTyrThrGly 2788
QY 10483 AAGAAATTAATGAT 10542
Db 2789 GluIleLysAsnTyrAsnValAsnMet---ValAsnThrMetAspLleProLleAsn 2807

```

QY 10543 GCTAAGAAAGTCTTAAAGAGCTAGTCTTAAATGAATGATCTAGT-GGTGCTAA 10601
Db 2808 ArgaspanasnavValtyrSerCylleaspleuileasnasPalaeuansgly----- 2825
QY 10602 CCTATTGATATATATGATGAACATATTG 10628
Db 2826 AspyrAspileTyfAspPoluValleu 2834
|||||
RESULT 6
Q26030 ID Q26030 PRELIMINARY; PRT; 3026 AA.
AC Q26030;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Variant surface protein (Fragment).
GN VAR.
OS Plasmodium falciparum.
NC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;
[1]
SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RX MEDLINE=95330813; PubMed=760678;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Welles T.E.;
RT "Type large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes.";
RL Cell 82:89-100(1995).
RN [2]
SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.D., Chittis C.E., Craig A.G., Roberts D.J.,
RA Hudson-Taylor D.E., Peterson D.S., Pinches R., Newbold C.I.,
RA Miller L.H.;
RT "Switches in the expression of Plasmodium falciparum var genes
RT correlate with changes in antigenic and cytoadherent phenotypes of
RT infected erythrocytes.";
RL Cell 0:0-0(1995).
RN [3]
SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RX MEDLINE=9094502;
RA Smith J.D., Kyes S., Craig A.G., Fagan T., Hudson-Taylor D.,
RA Miller L.H., Baruch D.I., Newbold C.I.;
RT "Analysis of adhesive domains from the A4VAR Plasmodium falciparum
RT erythrocyte membrane protein-1 identifies a CD36 binding domain.";
RL Mol. Biochem. Parasitol. 97:133-148(1998).
RN [4]
SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.;
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Kyes S., Smith J.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; U42244; AAD0351.1; -.
DR InterPro; IPR004258; PEMP.
DR Pfam; PF03011; PEMP; 2.
FT NON TER 3026 3026
SQ SEQUENCE 3026 AA; 348124 MW; EC0A971A82193788 CRC64;

Alignment Scores:
Pred. NO.: 3,29e-194 Length: 3026
Score: 4010.00 Matches: 1095
Percent Similarity: 42.80% Conservative: 480
Best Local Similarity: 29.76% Mismatches: 1151
Query Match: 20.61% Indels: 954
DB: 5 Gaps: 127

```

[illegible]


```

Db 1023 -----glylulyslleasnCysgly-----AlathrProtyrAspPtyrIle 1037
QY 3142 CCACAAAAATTAAGATGATGACCGCAATGGGACAAATGTACTGCAAGGTGCACAAAAG 3201
Db 1038 ProGlnArgLeuArgTyrMetThrGluTrpAlaGluTrpPheCysLysGluGlnSerArg 1057
QY 3202 GAGATATGATAGTTGAAGGAGCAAGTGTAGAGAGTGTAAAGATTAAGTATGTCAAAGC 3261
Db 1058 LeuTyrAspGluLeuMetGlyLysGlyIleCysMetAsn-----Gly 1072
QY 3262 TGTCAGAAAGAGAGTGTACAGGTGTACAGAGTGTACAGAGCTGTATGATATATAT 3321
Db 1073 IleCysAsnLysValLysAspAspCysAlaLysCysThrGluAlaCysLysGluTyrLys 1092
QY 3322 GATATATAGATATGATGACAAAGCAATGCAATATATATATATATATATATATATATAT 3375
Db 1093 ThrLysIleGlnProTyrLysAspGlnTrpGluLysLeuGluLysGluTyrAlaLeuSer 1112
QY 3376 GAATTCATATGACAAACGACAAATGTCTGTATAGTATAGTGTATGAAAGCTCCAGTACT 3435
Db 1113 TyrLeuHis----- 1115
QY 3436 GCCAAATATCATATATAGACAG-----AATGTT 3462
Db 1116 AlaLysAsnAspSerArgTyrMetAlaPheGlyGlyLysThrAspProAspTyrGlnGlnVal 1135
QY 3463 ATTGAATTTTGTGCGAATTTATATACCAACA-----AATGTGCGAAAGTATATATATATAT 3519
Db 1136 ValHisPhePheLysGluLeuGlnGlnAlaIleLysSerSerThrSerLysArgProLys 1155
QY 3520 ACTAGTATGATAAGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 3579
Db 1156 ArgSerThrPheAlaIleThrThrAspProThrThrProTyrSerThrAlaAlaGlyTyr 1175
QY 3580 CTCATGATATACAGAAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATG 3639
Db 1176 IleHisGlnGluIleGlyAsnAlaGlyCysGlnIleGlnLysHisPheCysAspAspAsn 1195
QY 3640 AGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3699
Db 1196 LysAsp-----LysTyrValPheArgGluLysProLysAspHisAspGlu 1210
QY 3700 GCGTGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3759
Db 1211 AlaLysAsnGlyThrGluAsnValGlu-----LysProLys 1222
QY 3760 GCGGAGAAAAAGATACGGA-----TGTAACAGATGATGATATCTTAAGAA 3810
Db 1223 LysGlnGluIleAspThrLysLeuAspValCysAlaIleValLysThrAlaLeuThrThr 1242
QY 3811 AACGATGAAAAAGAAACAGTAGAGATGTCATCCAAAAAAGAAATAGTATGATATCC 3870
Db 1243 GlnAspAsnLeuThrLysAla-----CysGlnGlnLys-----TyrGlyHisPro 1257
QY 3871 GAT-----TGCAATGC-----GGAATATATATATATATATATATATATATATAT 3909
Db 1258 GlnArgHisThrGlyLysPheLysCysIleSerGluThrThrThrLysSerSerAspSerGly 1277
QY 3910 CGTGTGTATGCGCCCTAGAACAAAGTATGCTGATCTTCTTGGCAAT----- 3963
Db 1278 SerIleCysValProProGlnArgArgLysLeuTyrValThrProLeuThrLysTrpAla 1297
QY 3963 ----- 3963
Db 1298 GluGluAlaThrGluGlyProThrSerProGlnAlaGlyGlyGluAlaThrLeuProThr 1317
QY 3964 -----GATATGAAATATATATATATATATATATATATATATATATATATATATAT 4011
Db 1318 ProAlaThrAlaSerSerGlnAlaProLysGlyAspSerLeu-----LeuLeuThrAla 1335
QY 4012 TTCATCAATGTGACAGACGAAACATTTCTTCATGATGATATATATATATATATATATAT 4071

```

```

Db 1336 PheIleGlnSerAlaAlaValGluThrPhePheLeuThrPheLysTyrLys---MetAsp 1354
QY 4072 GGTGAGAAATGCACTCGATTAAGAAATTAAGAAAGCAAAATTCCTCCGCAATTTTGG 4131
Db 1355 AsnAsnGlyLysArgLysAlaGlnAspLysLeuLysLysGlyGluIleProGlnGluPheLys 1374
QY 4132 AGATCCATGTTCTTCACATTTTGAGATTTATAGCAATTTTATTTTATTTGCAACA----- 4182
Db 1375 ArgGlnMetPheTyrThrLeuAlaAspTyrArgAspIleCysIleGlyValLysGluAsp 1394
QY 4183 -----GATATATCAAAAGGTGATGCTGAGGAGTGAACATTAACATTAAGAG--- 4224
Db 1395 ValIleLysAlaLeuGluAlaSerSerAspAsnLysSerGlyAsnAlaIleLysGluIle 1414
QY 4225 -----CAATAGATCTCTTTTCAAAAATGTCAGCAAAAATTCCTCAATGGA----- 4272
Db 1415 SerAspLysIleGluAlaIleLeuLysGlnSerGlySerLysLeuProGlyGluLeuPro 1434
QY 4273 -----AAACACGCCCAAGATGCTGACAGACAAACATGATGATGAGATA 4314
Db 1435 ValThrProAsnAsnValLysAsnProLysThrThrPheAspGlnAsnAlaLysHisIle 1454
QY 4315 TGGGAGCTATGCTATGCTGCTAGTGA-----AAATTTGGGCGCAAAA---AAA 4359
Db 1455 TrpHisGlyMetValCysAlaLeuThrTyrLysGluAspThrGlyGluAlaLysGlyLys 1474
QY 4360 GATGATTTTACCGAAGAAC----- 4377
Db 1475 ThrSerIleThrGlnAspProThrAlaTyrGlyLysLeuThrPheAsnAspGlyLysLys 1494
QY 4378 -----TACGTTACACACAGCTCAATTTAGT----- 4404
Db 1495 ProLysGluAspLysTyrAspTyrLysIleValLysIleSerSerValProSerSerAsn 1514
QY 4405 -----GACAAACACACACACTTGTGAGAAATTTGCAAAACACCCCACTTTTTCAGATGG 4458
Db 1515 ValProSerGlyAspThrLysLeuGlnLysPheSerArgArgProThrPhePheArgTrp 1534
QY 4459 CTAAACGAATGATGACGACGACTATGCTATACACGACAAATAATTTGAAAGATGTGACG 4518
Db 1535 LeuGlnGluTyrGlyGlnGluPheCysArgLysArgLysAspLysLeuGluLysIleGlu 1554
QY 4519 GAAAAATGTAG-----TCAATGACCAATTG 4545
Db 1555 LysGlnCysTyrGlyLeuAsnTyrArgLysArgLysIleTyrCysSerGlyAspGlyHis 1574
QY 4546 AAGTGTGAT-----ACAGAA 4560
Db 1575 IleCysGluLysThrAspThrSerArgAsnAsnThrPheIleAspLeuHisCysProArg 1594
QY 4561 TGTAATAGAAATGCGAGGACTACGTTAAATATATG-----AAAAAAGAGTGTGATTT 4617
Db 1595 CysLeuLysGlnCysIleLysTyrLysArgTrpIleGluLysLysGluLysGluPheHis 1614
QY 4618 CCACAAAGTAAATTTACAAAGATGAAACGACAAACAAAGATTCGATGACAAACATTT 4677
Db 1615 AsnGlnLysAsnAsnTyrGluLysGluPheAsnAsn-----AspLeuLysGluLys 1631
QY 4678 GGTGTAATGCTTACAGACTATACGTAACGAAACAGATCTTCTGACAGAAATTT 4737
Db 1632 Gly-----TyrSerSerPheAsn-----AsnPhe 1659
QY 4738 ACTGCTAGTTGTGTGAT---AACCTGGAAGTCCCTGTGTGACAAAGAAAT---ATA 4791
Db 1640 LeuAlaSerLeuAsnHisCysLysHisGlyLysHisAsnArgAspLysAsnLysIle 1659
QY 4792 CAATTGTAGAAAAACAGGCTTACTATGATCCGACAAACATTTGT---GGTGCACAAA 4848
Db 1660 GluPheAsnAsnHisThrLysThrPheGlyProSerGluTyrCysLysAlaCysProVal 1679
QY 4849 TTATTTGCAAAATGACGACAAATATATCTAACATTTTCAAGTAAAGATTAAGAAAGATTA 4908
Db 1680 TyrGlyValLysCysAsnLys-----LysAsnGlyGluCysGluThrIle 1694

```



```

QY 9103 TCTGAAAAATTTAAAAATATTTGAAACATCAATGAGCAACGAAATCGTAAACA 9162
Db 2817 -----
QY 9163 TGTGGGAAAAATATAGACGTCAGATATGCAACGCTATGTTATGATATATAATGCT 9222
Db 2817 -----
QY 9223 ACTTCAAAAGTAACTTAGATGAAGATGTCATTAATTCACCAAGATGAGAAAGCTAAT 9282
Db 2818 -----
QY 9283 CAGTTTCCTGCTGTTATATGATGGCAACAGCAACATGTAAGAAACAAACATGTA 9342
Db 2821 -----
QY 9343 AGTGATTCATTAACAAATGCTCGTTCAAACGAGATATTTGAAGCGTCAGAA 9402
Db 2822 -----
QY 9403 TTATTAGACACCTGATGTCAGATGATATTAAGAAATATATGCTTAATATATG 9462
Db 2825 -----
QY 9463 ATAAAAATATCAATGAAATCTAATATATAATATATAGCAATTAAGATCAATCTCA 9522
Db 2835 ValLeuHisProPheIleAsnMet-----
QY 9523 GGTATATATAGACAATTAACCATCTGAGAAATGTTCAATATATTAATCAAAAGAT 9582
Db 2842 -----
QY 9583 TCTCAATGCGCTTGGAGTTAATGATATTAATGAAATGTTACAGAAACAAATATAT 9642
Db 2843 -----
QY 9643 GAAAAATATGAAATTCAAAGAGTACTAAATAATATATATCTCGTTATATTTGTGAA 9702
Db 2860 AlaValProGlnPhe-----
QY 9703 GATGAACACACAAATAATCATGCTACTAGATGCAATATATAAAGAGAGCAACAGTT 9762
Db 2865 -----
QY 9763 CGTCTAAGCACTATTTCTTTACACCCAGTATGATTCTTCTATCAACACCTTCA 9822
Db 2874 AlaProPro-----
QY 9823 TTTCTCAACATCGAGTACACAAATATGATCTTAAATATATATGAAAAATAGTATC 9882
Db 2881 -----
QY 9883 TCTGTTGTTATGTAATGCGCTTAGTTGATAGCGCTTCATTCATTAAGAAAAATTC 9942
Db 2890 ProValIleGlyIleAlaIleLeuAlaIleLeuSerIleAlaPheLeuLeuLysLysThr 2909
QY 9943 AAATGCTGTGAGACTTGTGGTACTGATGATATCCCGAAGAGAGATATGAGATGCT 10002
Db 2910 LysHisProValAspLeuPheSerValIleAsnIleProLysSerAspTyrAspIlePro 2929
QY 10003 ACGTTGGAATCCAAAAATAGTATACATATACATATAGAGTCCATATTAAGCAAAA 10059
Db 2930 ThrLysLeuSerProAsnArgTyrIleProTyrThrSerGlyLysTyrArgGlyAsnGly 2949
QY 10060 ACATATATATATATATGAA-----GAGATATCTAGTGTGAGATGAAAGTAAATATATGCG 10113
Db 2950 ThrPheThrLeuLysGluIleValGlyThrAspSerGlyTyrThrAspHisTyr----- 2967
QY 10114 GACTTATCTCTCTGATATATCT-----TCATCGAAAGAGATGATGAAGATGATATAT 10170
Db 2968 -----SerAspIleThrSerSerSerGluSerGluTyrGluGluLeuAspIle 2983

```

```

QY 10171 AATGATATATATGTAACAGTACTCCTAAATATATAAATGATGATAGTACTAGAA 10230
Db 2984 AsnAspIleTyrHisHisValLeuLeuAsnIleLysThrLeuIleGluValLeuGlu 3003
QY 10231 CCATCAAAAAAGG-----GATATACCAAGAT-----GATACCAAGATAT 10272
Db 3004 ProSerGlyLysLeuSerLysSerLysAsnThrIleProThrSerGlyLysAsnThrProSerAsp 3023

RESULT 7
ID Q26031 PRELIMINARY; PRT: 3078 AA.
AC Q26031;
DT 01-NOV-1996 (TREMBLREL. 01, Created)
DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLREL. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR-1.
OS Plasmidium falci parum (isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmidium falci parum-
RL Cell 82:89-100(1995).
DR EMBL; L40608; AAA75396.1;
DR InterPro; IPR004258; PEMP.
DR Pfam; PF03011; PEMP; 2.
SQ SEQUENCE 3078 AA: 349297 MW: C8037C2BC3CDD7C3 CRC64;

Alignment Scores:
Pred. NO.: 1.36e-168 Length: 3078
Score: 3501.00 Matches: 1057
Percent Similarity: 38.65% Conserved: 470
Best Local Similarity: 26.75% Mismatches: 934
Query Match: 18.00% Indels: 1490
Gaps: 147

US-10-087-013-1 (1-10628) x Q26031 (1-3078)
QY 25 ATATATAAATGAGGATGCGAGATCATCATCTAGAGGAGATCTAAACCCCTATATA 84
Db 4 LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspIleGluLysP----- 20
QY 85 AAAGAAAGTCAAAAAGTGCAGAAATGTTTGGAACTGATGCCAAATATATAGACAT 144
Db 21 -----GluSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 36
QY 145 CCATCAAAA-----TATGCAAAAAGCAATGTCGATTCCTTAAGAGGGATTTGACGAAGCA 201
Db 37 LysValLysGluGluAlaLysGluArgGlyLysGlyLeuGlnIleArgLeuSerGluAla 56
QY 202 GAATTCGTGGTGGTCTCTACGCCAGTAAATAGCATATATATATATATATATCAATATCA 261
Db 57 LysPheGluLysAsnGluSerAspProGlnThrProGluLysP-----Pro 71
QY 262 TGTATATTTAGATCATATAGCAATATATATATATATATATATATATATATATATATAT 321
Db 72 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrThrAsnValIleAsn 88
QY 322 CCTTGCATGTAAGCAACAAACGATTTGATGGAAGTGAAGATTCGAATGT-----GGA 378
Db 89 ProCysAlaAspArgSerAspValArgPheSerAspGluTyrGlyGlyGlnCysThrHis 108
QY 379 AATTAATATCGTAATTAATAAAGAAATATGATGCTATAGCTGTGCGCACCTAGAGAGA 438
Db 109 AsnArgIleLysAspSerGlnGlnLysAspAsnLysGlyAlaCysAlaProTyrArgArg 128

```

OY	439	CGACATATGTCGTATAAAACCTTGAAGACGCTCTAAATGATATAAATACCAAAATATTCAT	498
Db	129	LeuHisValLysAspGlnAsnLeuGluGlnIleGluProIleLysIleThrAsnThrHis	148
OY	499	GATTATTTGGGAATGACATGACATGTTACAGCAAAATACGAAGTGAAATCAATTGTTAAAT	558
Db	149	AsnLeuValAspValCysMetAlaIaLysPheGluGlnSerIleThrGlnAsp	168
OY	559	CATCAAC--CATAAAGCACT-----TCAGACGCTTGACTGCTCTTCA	600
Db	169	TyrProLysTyrGlnIaIaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuIa	188
OY	601	CGAAGTTTGCACATATATAGTGCATATTTGTAAGAGAAATGATATGTTTAAACCAATGTC	660
Db	189	ArgSerPheAlaAspIleGlyAspIleValaIaArgLysPheLysLeuGlyAsnPro	208
OY	661	CATCAC-----AAAGTAGAAGACGGGCTCCGACAGTTCCTCAAGAAATA	705
Db	209	GlnGluIleLysGlnaIaArgGlnLeuGlnAsnLeuLysThrIlePheGlyLysIle	228
OY	706	CATGAT-----GGAATGAGAGATGAGTAAAAAATGATTACATCTGATGAGATCT	756
Db	229	TyrGluLysLeuAsnGlyIaGlnIaIaArgTyrGlyAsnAsp-----ProGlu-----	244
OY	757	GGAATATATATTAATTAAGACAGCATGCTGGTGCATGTGCATATGAAATTAAGTATGGCA	816
Db	245	-----PhePheLysLeuAaArgLysAspTyrPThrAlaAsnArgLysIleValTyrLys	262
OY	817	GCTATTAACATGTGATGCATCATATTAATCGAATATTATTCATATCAATCAAGAAATATCA	876
Db	263	AlaIleThrCysAsnAla--TrrGlyAsnThrTyrPhe--HisAlaThrCysAsnArg	280
OY	877	CCATTTATTTTCAATCCTTAATCGCGCCATTAACAAAGAAAGGTTCTCAACATTTAGAT	936
Db	281	GlyLysIaIaThrLysGlyTyrCysaArgCysaAsnAspAspGlnValProThrTyrPheAsp	300
OY	937	TATGTCCTCATATTTAGCTTGCTGTCACAGAAATGGGGAAGACGTTTCCGAAAAAGA	996
Db	301	TyrValaProGlnIaIaLeuAaArgPheGlnGluIaIaLysPheCysArgLysLys	320
OY	997	AATTTAAATTAATGAAGAAGTCAAGACGCTCGTCCG-----AATGACAAACAA	1044
Db	321	AsnLysLysIleLysAspValLysAlaGlnCysArgGlyLysAspLysGluAspLysAsp	340
OY	1045	CGCTTATATTTAGTACTCATATGACATGATTTAGCACACACATTTTGGAAAAAGATATT	1104
Db	341	Arg---TyrCysSerAaArgAsnGlyTyrAspCysGluLysThrLysaArgAlaIleLys	359
OY	1105	TTCGATTTGGATTAATTAAGTACTAGCTGCTGCTGACTAAATGCAAAGTTTTCAGACTTGG	1164
Db	360	LeuAaTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysaAsnProTyrValaAspTrr	379
OY	1165	TTTAGGATCAACAGACAGCATTTTAAAAAAACAAGAAATATGAAAAAGAAATATACA	1224
Db	380	IleAsnAsnGlnLysGlnIaPheAspLysGlnLysLysLysTyrAspGluIaLys	399
OY	1225	TCATAT-----TTATCAACAGATTAACAAATTTGTCAT	1257
Db	400	LysTyrGluAsnGlyAlaSerGlyLysaTrrGlnLysaArgAspAlaGlyLysIaThr	419
OY	1258	AATTTAATAGTGA--TATTTAAACAATTTTATGAAAAAATTAGAAACGCATAT	1314
Db	420	ThrThrAsnTyrAspGlyTyrGluLysLysPheTyrAspGluIaLysSerGlnTyr	439
OY	1315	GCACACTATGACACTTTTAAATTTATTAACATTAAGAAAGAAAGATATTCG-----	1362
Db	440	ArgThrValaAspLysPheLeuGluLysLeuSerAsnGluGlnIleLysThrLysValLys	459
OY	1363	-----AAAGAGGATTAACAGAGAAAGAAAGATATTACTTTTACTACAGTCT-----	1410
Db	460	AspGluGlnGlyLysThrIleAspPheLysAsnValaAsnSerAaSerThrSerGlyAla	479
OY	1411	-----GATGACAAAGGAGATATTTTATCGTTCAGAAATATTCGCCAAGTGTGT	1455

Dd	480	SerGlyThrAsnValGluSerGlnGlyThrPheTyArgSerLysTyCysGlnProCys	499
Qy	1456	CCCGACGCGCGGGGTCAAAATGTCATGGTAAATATACACAC- :::	1512
Dd	500	ProTyCysGlnValLys	515
Qy	1513	CGTAACTGTAATAAT	1554
Dd	516	TPrpLglLysAsnGlnGlyLysCysLysSerGlyLgLeuTyGlnProLysProAsp	535
Qy	1555	GTGAAGCGTACTATATACACTGTCCTTTATAGCGTAAAGAACAAAGGATATTACAA	1614
Dd	536	LysLgLeuGlyThrThrIleThrIleLeuLysSerGlyLysGlnAspAspIleGlu	555
Qy	1615	AAATTGAAAAAATTTTGTAACTAC	1641
Dd	556	LysLeuAsnLysPheCysAspGlnLysAsnGlyAspThrIleAsnSerGlyGlySerGly	575
Qy	1642	ACTAATTACAAAGATAAAAATAAT	1680
Dd	576	ThrGlySerGlySerGlyGlyAsnSerGlyArgGlnGluLeuTyGlnGluIleProLysCys	595
Qy	1661	TATTATAGGATGCAAAATATTAATAGATGTGAACCTGGACAAATAATCTGAAATCAATAT	1740
Dd	596	--TyrlLysGlnLysPheValValLysValGlnHisAspGlnAspAspGlnGluAspTyr	614
Qy	1741	GATATCTCTTAC	1758
Dd	615	GlnAsnValLysAsnAlaGlyGlyLeuCysIleLeuLysAsnGlnLysLysAsnLysGlu	634
Qy	1759	-----TCAATTCATTAATTT	1773
Dd	635	GlnGlyLysAsnThrSerGluLysGlnProAspGlnIleGlnLysThrPheAsnProPhe	654
Qy	1774	TTTGAATTATGGGTACATATTTATTATAGGAGTACTATTAACTGGATGACAACTTAA	1833
Dd	655	PheTyTrpTyrValAlaHisMetLeuLysAspSerIleHisTrpLysLysLeuGln	674
Qy	1834	ACTGTATA--AAATATATACACACAGCATGT--ATTGATGAATGATACAGAAATGTC	1887
Dd	675	ArgCysLeuGlnAsnGlnLysAsnAlaGlyIleLysCysGlyLysAsnLysCysAsnAsnAspCys	694
Qy	1888	TTATGTTTGCACAGATGGTTAAACAAAGAAAGAAATGATATTAAGAAACTG	1947
Dd	695	GlnCysPheLysArgTyrIleThrGlnLysLysAspGlnIleProGlyLysIleValGlnHis	714
Qy	1948	TTTCACAAAAAAGATATACAG	1971
Dd	715	Phe--LysThrGlnAsnIleLysGlyArgGlySerAspAsnThrAlaGluLeuIle	733
Qy	1972	-----CAATGATTTATAGTAATATTAATATCTTTTGAGAGTTATTT	2016
Dd	734	ProPheAspHisAspTyrValLeuGlnTyrAsnLeuGlnGlnLysPheLeuLysGlyAsp	753
Qy	2017	-----TTTAAAGTTATGATATTAATCTTGACAAAGATGACCAAAATGG	2058
Dd	754	SerGlnAspAlaSerGlnGluLysSerGlnLysSerLeuAspAlaGlnGluAla	771
Qy	2059	AAAGACTTATGAAAAATATTAAGAAAAAATAATGACTTTTCCAAATTTGGAAATAAT	2118
Dd	772	-----GlnGluLeuLysHisLeuArgGluIleIleGluSerGlnAspAsn	787
Qy	2119	AGGGAC-----TATTAGCAATGCAATAGCAACTCTTGTA	2154
Dd	788	GlnGlnAlaSerValGlyGlyGlyValThrGlnGlnLysAsnIleMetAspLysLeuLeu	807
Qy	2155	GATCACTTAAAGAACTGCCAGCATATGT-----	2184
Dd	808	AsnTyrgluLysAspGlnValAspLeuCysLeuGlnIleHisGlnAspGlnGluGln	827
Qy	2185	---AAAGCATATATACAAACGAAGCATGTGAACATCCCATATGCAACAAACCCG	2241

Db	828	LysGlnLysGlnLysArgLysAsnGlnLysSerLleGlnGlnGlnLysLysAsnPro	847
QY	2242	TGCGTTAAACCTCGGAGAGCCAA-----CCACATAAAATTAAGAATATGACACA	2298
Db	848	CysSerGlnLysSerGlnLysLysAlaGlyTyrProValLeuAlaLysValAlaTyrGln	867
QY	2299	TACTTTAAAGAGGATCATGACGAGGAAGACAGAAATCGATGCTTCATTAATTGAAGA	2358
Db	868	MetHisHisLysAlaLysThrGlnLeuAlaSerArgAlaLysArgSerAlaLeuArgGly	887
QY	2359	AAGGACACAGGAAGATATATATTAACACCTGGGGGTAGAGAAAGGACTTCAGCAAAATTA	2418
Db	888	AspIleSerLeuAlaGlnPheLysAsnGlnLysArgAsnGlySerThrLeuLysGlnIle	907
QY	2419	TGTAGATATATGATTAACATTTTAAACGTAACCTGCTTTTGAATTCGACCATGAT	2478
Db	908	CysLysLleAsnGlnLysTyrSerAsnAspSerArgGlnLysSerGlyLysProCysThr	927
QY	2479	GCCAAAGCACAGGTGATGATACAAACAGATTTGTGTGAGAACTGAATGGGAAAGTG	2538
	928	GlyLys---AspArgPheHisGlnGlyValArgMetArgLleGlyThrGlnLysSerAsn	946
Db	2539	GATCCGGAACACATGCCGTAAGATACAGAGATGTATATGCTCTCTGGAAGACGAT	2598
QY	947	LleGlnGlyLysLysGlnThrSerTyrLysAsnValPheLeuProArgTyrGlnHis	966
Db	2599	ATATGTATCATCCAAATTTGGAACATTTTACAACGATGATACCCACTTAATGGTAAT	2658
Db	967	MetCysThrSerAsnLeuGlnLysAsnLysVal-----GlySerVal	980
QY	2659	GTT---GATGATTTAGTTAATTAATTAATCTTTTGGGGAGTCTTCTATACGAAATAT	2715
Db	981	ThrLysAsnAspLysLysAlaSerHisSerLeuIleGlnLysValGlnLeuAlaLysThr	1000
QY	2716	GAAGCAAAACAGATATATACGATATGAAGAAAGAAATAACTTAAGGCCCCAAAGAA	2775
Db	1001	AspAlaLysAlaGlnLleLleLysArgTyrLysArgGlnAsnAlaIle-----Gln	1016
QY	2776	GTAAGTACCCCA-----AAACACACACAACTATCTGTGAGCTATACGTTACAGT	2826
Db	1017	LeuThrAspProLleGlnGlnLysAspGlnGlnLysAlaMetCysArgAlaValAlaGlySer	1036
QY	2827	TTTGACATATATAGGTATATTAATTTCCGGAGAGAGATCTGTGGAGAAAGAAAGGATGACATG	2886
Db	1037	PheAlaAspLeuGlnLysArgLleIleAlaGlnGlyArgAspMetThrAspGlnLysSerSer	1056
QY	2887	GTAAGCTGCAGGACATTTTGGAAACTGTTTTGTATATATACATTAAGTCACTCAAAAGC	2946
	1057	ThrAspMetGlnThrArgLeuIleThrValPheLysAsnLleLysGlnLysHisAspGly	1076
Db	2947	--AAGGAATGATTAATTAATTAATGATGAT-----GCCCCCAATATTTAAATTG	2994
Db	1077	LleLysAspAsnProLysTyrThrGlnLysAspGlnSerLysProAlaTyrLysLysLeu	1096
QY	2995	AGGAAATATGTTGGGGAAGCTAATATAGACCAAGATATGGGAAGCCATGAATATGATATA	3054
Db	1097	ArgLysAspTyrPThrGlnLysAlaAsnArgHisGlnValTyrArgAlaMetLysCys-----	1114
QY	3055	AAATATTTGAAGGATTAATTCGGGACACCAATGACAACAAAGATAGTATTCGGCATATAGT	3114
Db	1115	-----AlaThrLysGlnLleIleCys-----	1121
QY	3115	GATCATACACCATTTGGATGTATATATCCACAAAAATTAAGATGATGACGATGAGGCA	3174
Db	1122	ProGlnMetProValAspAspTyrLleProGlnArgLeuAlaGlyThrMetThrGlnTyrAla	1141
QY	3175	GAATGTACTTCGCAAGTGCAGAAAAAGAGATATGAATGAAGGAGAAAGTGAAGAG	3234
Db	1142	GlnTyrPyrCysLysLysAlaGlnSerGlnGlnTyrAspLysLysLysLysLleCysAlaAsp	1161
QY	3235	TGTAAAGATTAAGGATTAATGTCGTAAGGCTTCGCAAAAGAGATCGTACAGTTGTACAGAG	3294
Db	1162	CysMetSerLysLysGlnLysLysLys---CysThr---GlnGlnLysValAspCysGlnLys	1179

OY	3295	TGCACGAAAGCGAATGAATTAATGATAATGATTAGCAAAACAATGAAAT	3355
Dd	1180	CysLysAlaIleCysAspLysTyrLysGlnGluIleGluLysTrpSniGlnIleTrpArg	11995
OY	3355	ATAATATCAGATAAATACMAAGAATTTCATGACACAGCACAAATGGCTGTAGTAAGT	3414
Dd	1200	LysIleSerAspLysTyrAsnLeuLeuTyrLeuGlnAlaIleValThrThrSerThrHisPro	12119
OY	3415	GGTATTGAAGCTCCAGACTGCCAAAATCATATAGAC-----AGCAATGTAT	34655
Dd	1220	Gly-----ArgThyValLeuGlyAspAspAspProAspArgGlnIleMetVal	12335
OY	3466	GAATTTTGTCGAATTATACCAAAATGGTGGCAA-----AGTAATAAA	35135
Dd	1236	AspPheLeuThrProIleHisLysAlaSerIleAlaAlaArgValLeuValLysArgAla	12555
OY	3514	ACTGGTACTAGTAGAAGAAAGCTGCTCATTTGGTACTAACACACGATGAAAAATTGGA	35737
Dd	1256	AlaGlySerProThrGlnIleAlaAlaAlaProIleThrProTyrSerThrAlaIle	12757
OY	3574	GCATGTCCCATGATACAGGAATTTTGATGATGTGCATCCAAATAGACTTTGTGAT	36333
Dd	1276	GlyTyrIleHisGlnGluIleGlyTyrGlyGlyCysGlnGlnGlnIleGlnPheCysGlu	12955
OY	3634	GAATA-----AGTAGGTAAAGATGAAATAATCCCTTTAGA	36757
Dd	1296	LysLysHisGlyAlaThrSerThrSerThrThrLysGluAsnLysGluTyrThrPheLys	13155
OY	3676	GATTAACACAGAGCACTGATGGTGGCGTGCGTGTGTAAAGTGATCGAAACCACAAAG	37355
Dd	1316	GlnProProProGlnTyrAlaThrAlaCysAspCysIleAsnAspSerGlnIleGlu--	13344
OY	3736	GTACAGATAAAAAGCAAAAAAGCGGAAGAAAGATACGGAATGTAAAAAGTAGAT	37955
Dd	1335	--GlnProLysLysGlnGluAsnValGlnLysAlaCysLysIleValGlu	13511
OY	3796	GATTACTTAAAGAAACGATGGAAGAAACAAGTAAGAAATGTTCATCAAAAAAGAT	38555
Dd	1352	LysIleLeuGlnGlyLysAsnGlnGlyArgThrThrValGlyGlnCysAsnProLysGluSer	13711
OY	3856	AGTAATGGAATCCCCGATGGCAATGC---GGAAATATATATTTGTGGAAGCCCTCGT	39112
Dd	1372	-----TyrProAspTrpAspCysLysAsnAsnIleAspIleSerHisAsp---Gly	13877
OY	3913	GTGTGTATGCCCTACAGACAAAGATTATGCGTACATTTCTTGCCAATGATATAGAA	39712
Dd	1388	AlaCysMetProProArgArgGlnLysLeuCysLeuTyrTyrIleAlaHisGlnSerGln	14077
OY	3973	ATTAATAAATTTACATTCACAAAGTATTATTAAGAAAGCTTCATCATATCTGCKAGCA	40312
Dd	1408	ThrGluAsnIleLysThrAspAspAsnLeuLysAspAlaPheIleLysThrAlaAlaIle	14277
OY	4033	GAACATTTCTTCATAGTATTATTATTAAGTAG--GATGGTGAAGGAATGAACTC	40897
Dd	1428	GlnThrPheLeuSerTrpGlnTyrTyrLysSerLysAsnAspSerGlnAlaLysIleLeu	14477
OY	4090	GATTAACAATTAAGAGCAAAATTCCTCCCGCATTTTGGATGTCATGTTCTACACA	41497
Dd	1448	AspArg-----GlyLeuIleProSerGlnPheLeuAlaGlySerMetIleTyrThr	14637
OY	4150	TTTGAGATTATAGATTTTATTATTTGGAACAGATATATCAAAAGGTCAAGTGAGGGA	42097
Dd	1464	PheGlyAspLysArgAspIleCysLeuAsnThrAspIleSerLysLysGlnAspVal	14837
OY	4210	AGTAACTAAAGACAAATAGATTCTCTTTCAAAAATGGTGCAAAAATCTCTAAT	42697
Dd	1484	AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspGlySerLysSerProSer	15037
OY	4270	GGA AAAACAGCCAAATAGTGGACAGAAACATAGTCACTGAGATATGGAACTATGCTA	43297
Dd	1504	GlyLeuSerArgGlnIleUtrPirPyrLysThrAsnGlyProGluIleTrpLysGlyMetIle	15237

QY 4330 TGTGCACTAGTAAATTTGGGCAAAAAAGATGAT-----TTTACCGAAACTAC 4380
DB 1524 CysAlaLeuThrIysTyrValIthrAspAsnIysArgLysIleIysAsnAspTyr 1543
QY 4381 GGTTCACACAGCTCAAAATTTAGTGACAAA---AGCACACCTTTGGAGAAATTTGCCAAA 4437
DB 1544 SerTyrAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnIleuPheAlaIle 1563
QY 4438 CGACCCCACTTTTACAGTGGTAAACCGAATGGTACGACACTATTGCTATACACGACA 4497
DB 1564 LysProGlnPheLeuArgTyrMetIleGluTyrGluGluPheCysAlaGluArgGln 1583
QY 4498 AATATATTAGAGATGCGAGGAAAAATGTAAGTCA---AATGACCAATTTGAAGTGTGAT 4554
DB 1584 LysLysGluAsnIleIleLysAspAlaCysAsnGluIleAsnSerThrGlnGlnCysAsn 1603
QY 4555 ACA-----GAATGTATATAGAAATGCGAGACTAGCTTAAATATATGAAA---AAA 4602
DB 1604 AspAlaLeuHisArgCysAsnGlnAlaCysArgAlaTyrGlnGluTyrValGluAsnLys 1623
QY 4603 AAAAAAGCTGGATTTCCACAGATAAATATTACAGAGATGACGACAAAAAAGATTC 4662
DB 1624 LysLysGluPheSerGlyGlnThrAsnAsnPheValIleuLysAlaAsnValGlnProGln 1643
QY 4663 GATAGCAACACATTGGTGAATGTTAGACACTATCTGGAACGATGCACAGATTAC 4722
DB 1644 AspProGluTyrLysGlyTyrGluTyrLysAsp----- 1654
QY 4723 TTGACAGAAATTTACTGCTAGTTGTGGTATAGCTGGAAGTGCCTCTGTGTACAA 4782
DB 1655 -----GlyValGlnPro-----IleGln 1660
QY 4783 AGAATATATCAATGTTTAGAAAAACAGGCTTACTATGATCCGACAAACATTTGGGTGC 4842
DB 1661 GlyAsnGluTyrLeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1676
QY 4843 ACAAAATTTATTGAAATATACGCAATATACTACATTTCCGATTAAGTACAAA 4902
DB 1676 ----- 1676
QY 4903 GGATTAGTAAGAGCGCAACACAGGTGCTATTAACTGGCAAAACAGTCCATAATAC 4962
DB 1676 ----- 1676
QY 4963 TACATTAAGTGAAGAAATGACTGAAGATGTGCTTTTCTCTCTGCGACTAGCTATA 5022
DB 1676 ----- 1676
QY 5023 TGTTCATGCAATGATGCAATTTATACAGATCCAGAAATTAAGTAAAAATGGGTGC 5082
DB 1677 -----MetAspLysAsnValLeuSerValSerProLysGluLys----- 1689
QY 5083 CGAAAAAGATTGATGGAATGTGGCGCAACGAGAGGTTACAAATTTGGTCAATACTACAA 5142
DB 1690 -----ProPheGlyLysTyr----- 1694
QY 5143 GAAAAAAGAAAAAGAAAAATAAAAACGTGCGATCCGCACAAATATTTCTATGAGTTC 5202
DB 1695 -----AlaHisLysTyrPro----- 1699
QY 5203 CGGCCTTGTAGTCTATGAATATAGTTTTTATGATTAAAGATATTAATTTAGTATT 5262
DB 1699 ----- 1699
QY 5263 GATTAATTTGAGAGTAAAAACAAAAAGCCAGAGAAATTTGAAAGAAATATTTAACAA 5322
DB 1699 ----- 1699
QY 5323 AATGCAATCAGCTTGGCAAGAGAGATAGTACTACAGAAATCCCGTACTACTGCG 5382
DB 1699 ----- 1699
QY 5383 CGAAATTTTCTGGAACGAAAAATAGAAATGTGTGGAACCATGATATCGGGTAC 5442

DB 1699 ----- 1699
QY 5443 AAACGTGTAGGGATGATGGAATAGTGAATATAGTCAAGAGATGATGAAGATCTAAAA 5502
DB 1699 ----- 1699
QY 5503 AAATGTGTTCTGTACCTTCAGATGATGATTATCTATGGGAAAAATCGCGATGAAGT 5562
DB 1699 ----- 1699
QY 5563 ACTGCTATCAGTTTCTTCGATGTTTCCGAATGGGGTGAAGATTTTGCAAACATAAA 5622
DB 1699 ----- 1699
QY 5623 GAAAAAGAAATTTGGAAATTTGTAGGGCGGTGAATATTAATTAATTAATTAATGA 5682
DB 1699 ----- 1699
QY 5683 GATTAAGAAAGAAATGTACAGATGCGGTACACAAATATTAATTAATTAATTAATGA 5742
DB 1700 -----GluLysCys----- 1702
QY 5743 AAACACAGTATGAAAAACAAATCAAAAAATATGTGAGATTAAGCAAAATATATTC 5802
DB 1702 ----- 1702
QY 5803 GAGCATCTGTGCAAAAGATGACAGAGAGCTCGGAATATTTAGCAAAATTAATA 5862
DB 1702 ----- 1702
QY 5863 AAATTTTGAAAAATAAAGTGAATTTGAATATATAGTATGAAGATGTGTCCACA 5922
DB 1703 -----AspCys----- 1704
QY 5923 CAGCATTAACGTATGTTAATGTAATAATATATCCCGCATCATTAAGCATGAACAAA 5982
DB 1704 ----- 1704
QY 5983 GAAGTTGAAGAAAGTGAATATGTCAGAGTCCACAGAGTCCACAGCTGTACGAGGAA 6042
DB 1705 -----TyrGlnGlyLys-----HisValProSerIleProPro----- 1715
QY 6043 ACACGCTCACCGGGTATCTGATATCAAAAACGACGGCATGCAAAAAAGCGAAA 6102
DB 1715 ----- 1715
QY 6103 ACAGCCCGGCTCAAAAACAGCCGAAAAAGTGAATATTAACACAGAAATGCGAGCA 6162
DB 1716 -----ProProProProValGlnProGln----- 1723
QY 6163 CAACACAGAACCCGACGACGACACAAACAGAAACGAAACATCAACAGCAACACA 6222
DB 1724 -----ProGluAlaProThr----- 1728
QY 6223 ACAGAACTGACGTGGGACCAATGTAAGGCCATTTCTGAATAAACAGATACAGG 6282
DB 1729 ValThrValAspValCysSerIleValLysThrLeuPheLys-----AspThrAsn 1745
QY 6283 GTGGAATAGAGGTTGTAATCCAAAAACGTATAGGACAAAT-----CTTAATGGGGT 6336
DB 1746 AsnPheSerAspAlaCysGlyLeuLys---TyrGlyLysThrAlaProSerSerTyrLys 1764
QY 6337 TGTATT-----GTAGTAACTTAAGAAAT 6363
DB 1765 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrLysSerGlySerAsp 1784
QY 6364 GAAATGCAATATGATGCTCTAGAGAAAAAATATATGATTAATAATATATATATATAT 6423
DB 1785 SerGlySerIleCysIleProProArgArgArgLeuTyrValGlyLysLeuGln--- 1803
QY 6424 TTAATTAAGAACT-----GAAATTAAGCTGAC 6453

Db	1804	-GUTPRPAlaThrAlaLeuProGlnGlyLeuAlaIleProSerHisSerArgAla	1822
QY	6454	AATGATATAAAGAGCGCTTTATTAAATGTCAGCAAGATAGAACTCAATTGGTGTTA	6513
Db	1823	AspAspLeuArgAsnAlaPheIleGlnSerAlaIleGlnIuThrPhePheLeuTrpAsp	1842
QY	6514	AAATATATAATTGAA--AATCCGACAGCAAAAAATGAA-----	6549
Db	1843	ArgTyrLysGlnGluLysLysProGlnGlyAspGlySerGlnAlaIleSerGlnLeu	1862
QY	6550	-----TTGCAAAATAGGAACA	6564
Db	1863	ThrSerThrTyrSerAspAspGluGluAspProProAspLysLeuLeuGlnAsnGlyLys	1882
QY	6565	ATTCAGATGGAATTTAAAGAAATATGATATATACATATGGTGATTATTAAGATATGTTT	6624
Db	1883	IleProProSerPheLeuIleGlyMetPheThrIleGlnLysAspTyrArgAspIleLeu	1902
QY	6625	TTTGGAACTGATATTTCTAATGATGATAAAAAAATTTATACCTGACAAATAGTGTAACAAC	6684
Db	1903	ValHisGlyGlyAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle	1922
QY	6685	ATTCATCAATGAAATTAATAAGAAAACAGCATGATAAAAAAGATGAAATTAACGTAA	6744
Db	1923	ValLeuGluAlaSerGlyAsnLysGluAspMetGlnLysIleGlnGluLysIleGluGln	1942
QY	6745	ATA-----	6747
Db	1943	IleLeuProLysAsnGlnGlyThrProLeuValProLysSerSerAlaGlnThrProAsp	1962
QY	6748	-----TTTTGGCAGAAAAATAAAAATTTATTTGGCAAGCAATGATATATGATTAACCTAT	6804
Db	1963	LysTrpTrpAsnGlnHisAlaGlnSerIleLeuTrpGlyMetIleCysAlaLeuThrTyr	1982
QY	6805	-----CATCTCACAGACGAAACGAAAAAGAAAAATTTAGA	6840
Db	1983	ThrGlnLysAsnProAspThrSerAlaArgGlyAspGluAsnLysIleGluLysAspAsp	2002
QY	6841	GATAAATTAC-----	6849
Db	2003	GluValTyrGluLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro	2022
QY	6850	-----CAGTACAT--GACATGACCAAACTG-----	6873
Db	2023	ThrGlyThrTyrLysThrGlnIuThrAspTyrGluLysValLysLeuLysAspThrSerGly	2042
QY	6874	-----ACGCTTCCTCTGAAGAGTGTGTAATAAAGGCC	6906
Db	2043	AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheAlaLeuArgPro	2062
QY	6907	CATTTTGGATGGTGTTCCACAAATGGCAGAAAGATTTTGTAAATPAGAGAAAGAC	6966
Db	2063	ProTyrPheArgTyrLeuGlnIuThrTrpGlyGlnAsnPheCysLysLysArgLysHisLys	2083
QY	6967	TTTGTAAAAATTGGAGCGCGCTGTAAAGCAATATGAG-----	7002
Db	2083	LeuAlaGlnIleLysHisGlnLysLysValaGlnGluAsnGlyGlySerArgArgGly	2102
QY	7003	-----TGTAAAT-----GGTAAAT	7017
Db	2103	GlyIleThrArgGlnIuTyrSerGlyAspGlyAlaLysAsnGlnMetLeuProLysAsn	2122
QY	7018	GACGGTAGACACAAGAA-----TGCGAGAGCGCTGTGTAAACATATCA	7062
Db	2123	AspGlyThrValProAspLeuGlnLysProSerCysAlaLysProCysSerSerTyrArg	2142
QY	7063	AATTTATTAAGACGTGAAAACTGCAATATGAAACAAAGAAAGTTCAAAAAGGAT	7122
Db	2143	LysTrpIleGlnSerLysGlyLysGluLysPheGlnLysGlnGlnLysAlaTyrGluGlnGln	2162
QY	7123	AAAGAT-----GGCAAAAAGTATAGATTTCTTCTACGAAAGACATA	7170
Db	2163	LysAspLysCysValAsnGlySerAsnLysHisAsp-----	2174

OY	7171	GAGAGGCAACATGCTGCTCATGAAATATTTAAACATGAAATTTAAAGAAATATATGCGCAAT	7230
Db	2175	-----	Asn 2175
OY	7231	AAGGATGCTTCGTATGTCAGAAAAACCTTCACAACTACCAAAAAACAACAAACAAAT	7290
Db	2176	GLYPheCysGLuThrLeu-----	Thr 2182
OY	7291	CAATCATCCGATGCTAATGATATAGCCAAATCGTGATATATGTTCCGAAATTTAAC	7350
Db	2183	ThreSerIysAlaIysAspPheLeuIysThrLeuGly-----	2195
OY	7351	AAGTGTGCTGCTGAACTTCCTGAAAAAAGGAGCATATGAT	7392
Db	2196	-----ProCysIysProAsnAsnValGluLeuIysThrIlePheAspAspIysThr	2213
OY	7393	-----CATACAAAAAAATTTACTGAACT-----AAATATCCTATGAATGTGTAGAG	7440
Db	2214	PhelIysHisThrIysAsp--CysAspProCysLeuIysPheSerValAsnCys-----	2230
OY	7441	AAACGACGATATTTATATCTAAAGAACGAAATAATATGATATTACCTTGAAAGAA	7500
Db	2230	-----	2230
OY	7501	AAATTTATACCTATTGAGTCTACAAAGAAAGAAAGTAAAAATAGTTGGACATATAT	7560
Db	2231	-----LysLeuAspGluCysAspAsnSer-----	2238
OY	7561	AATCCTTGCGATCCTTAAGAAACCTTTATGACCTGATTAATATATAGGAAGAAAGAACCT	7620
Db	2239	-----LysGlyThrAsp-----	2242
OY	7621	TGTGAAATATAGAGAAATCGTTTAAAGTAGATATGAAATGAAATGTTCAAAAT	7680
Db	2243	CysAlaGsn-----	2247
OY	7681	TCAAAGTCTATATCAGAGAAAAAAAGAGTATGTACTCCAGAAAGAACATATATGTGC	7740
Db	2248	Set-----	2248
OY	7741	TTTAAAGAAATTTAGATGAAATTTAAATTTGAAAGACTTAAGATATGTAATATCTCCTAAAA	7800
Db	2249	-----IleAspIaIhrAspIleGluAsnGlyValAspSerThrValLeuGluMet	2265
OY	7801	ATGCTTCGTGCAACTCAGCAAAATGGAAGAAATAGACATATATAAAACTTCAACTCA	7857
Db	2266	ArgValIserAlaAspSerIysSer-----GlyPheAsnGlyAsp	2278
OY	7858	-----GAGAAAGGGGGCCCAATGAATCCAAATATGTGATACTATGAAATATATAGTTGGCT	7911
Db	2279	GlyLeuGluAsnAlaCys-----	2284
OY	7912	GATCTGGGTAGCATAGTATGTTAGAGAAACAGATATGTTACGAATGTGGTACTTACCTCC	7971
Db	2285	-----ArgGlyAlaGly-----	2288
OY	7972	GTTGAATATAATTTATATAGGTTTGTGAAATCAATATATGAAAAATGGAATAATAAAT	8031
Db	2289	-----IlePheGluGlyIle-----	2293
OY	8032	AAAGGTAGAAATAATAATACAAAGATGTACAAAGCTTTCGTTCTGCTTGGTGGATGCTAAT	8091
Db	2293	-----	2293
OY	8092	AGAAAAGATATTTGGAAGCA-----ATGACGTGCAAAAGCACCAGAA	8133
Db	2294	ArgIysAspGluThrIlyCysAlrAsnValCysGlyTrpValIvalCysLys--ProGlu	2312
OY	8134	GATGCAAAACTTTTAAAGAAAGAAATGATGATGATTGGAACGATACATTAATACAA	8193
Db	2313	AsnValAsnGlyGluAlaIysGlyLys-----	2321

QY 8194 GATPAGTGTGACATGAGACGATCCACTGTGATGATATATACCTCAACGTTTCA 8253
Db 2322 -----HisIleIleGlnIleArgAla 2328
QY 8254 TGGATGACGATGAGTGTGATATATATGTAAGCACTGATGAGAAATGGAATAATTT 8313
Db 2329 LeuValIysArgTyrValGluTyrPhe-----PheGluAspTyrAsnIle 2344
QY 8314 AAAAATCATGTGACTACTTAAACATCTGACAGATGACGAATGATTAATGAAAT 8373
Db 2345 LysHis-----LysIleSerHisArgIleLysAsn-----Gly 2355
QY 8374 AAGTGAACAGCTGTAAGACGAGATGTCAGAAATATATAAATTTGTTTAAATGAAA 8433
Db 2356 GluIleSerProCys-----IleLysAsnCysValGluIleTyrPhe 2368
QY 8434 TCTCTATTCGATATACATCAAAATTAATCAAGAAATGATGATGATGAAACCAATATATCA 8493
Db 2369 -----ValAspGlnIleValGluTyrPheGluIleThrGluArg----- 2382
QY 8494 AAAATCTCTACTTATGATCATGTTCAAAATTTGTACAAAAGTTGAAAACCTTTTAAAGT 8553
Db 2383 -----PheLysAsp 2385
QY 8554 GAATGTTCTGTGAGACCTTTTCTGAATATCTTCATGAAACAGTAAGTTTGAATAT 8613
Db 2386 Gln-----Tyr 2387
QY 8614 AAATTAATGAAAATGATGTTCTTCCATATATACGACATATGTTTCGAAGAACCA 8673
Db 2388 Lys-----AsnAspAsnSerAspAspAsnValArgSerPheLeuGluThrIlePro 2406
QY 8674 AAA-----ACTTATTAAGAAAGCT 8691
Db 2407 GlnIleThrAspAlaAsnAlaIleValIleLysLeuSerLysPheGlyAsnSer 2426
QY 8692 TGCAGTGTGACACTACTCTTCAAGATCCATGATGATATGTCTACCAGATCAAAACAA 8751
Db 2427 CysGluIleCysSerAlaSerAla-----AsnGluGlnAsnLys 2438
QY 8752 GATGAGATGTAAGCAATATACAACTTTACCTCTGTCGAGAAATGATTAATATATAT 8811
Db 2439 AsnGly----- 2440
QY 8812 CTGTGATTAATGGAACGATACCTGTTCTTAATAGTTGAGATGATTAACAAAGTGTATG 8871
Db 2440 ----- 2440
QY 8872 ATTCTCCAGAGAGACATTTATGACAGACCTATCACTGATATATATAGAAA 8931
Db 2440 ----- 2440
QY 8932 GGTGATTAAGCAATTTTAAAAAACTTTCTACTCTGCTTCAGTCAAGACAAATG 8991
Db 2440 ----- 2440
QY 8992 TTAGCTCAAAAAATTAATGGAAGAAAGTGTGCTTTGAGCAATGAATATATAGTTAT 9051
Db 2440 ----- 2440
QY 9052 GCAGATTAATCCGATATATTAAGAACTGATATGATGACACTTCATTTATGAAAA 9111
Db 2441 -----GluTyrIleLysAspAlaIle-----AspCysMetLeuLysLysLeuLysAspLys 2456
QY 9112 ATTAATAAAATATTTGAACATCAATGAAGCAACGAAATCTTAATAACATGTGTGGAA 9171
Db 2457 Ile----- 2457
QY 9172 AATTAATGACGTCAGATATGCGACGCTATATGTTATGATGATTAATAATGCTACTTCAAA 9231
Db 2457 ----- 2457
QY 9232 GTAACATTAGATGAAGATGTGTCAATTACCAAGATGAAGAACTAATCAGTTTCTT 9291

Db 2458 -----GlyGluCys----- 2460
QY 9292 CGTTGTTAATTAATGAAATGGCAACGATGTAAGAAAGAAACATGTAAGTATCA 9351
Db 2461 -----GluLysLysHisHisGlnThrSer 2468
QY 9352 TTAATAACAAATGT-----CCTGTTCAACGAAGATTAATTTGAAAGCG 9396
Db 2469 -----AspThrGluCysSerAspThrProGlnProGlnThrLeuGluAsp----- 2483
QY 9397 TCAGATTAATTAAGACAACTCGATGTCAGATGATATTAAGAAATATATAGCTTGAT 9456
Db 2484 -----GluThrLeu-----AspAspAspIleGluThrGluAlaLysLys 2497
QY 9457 ATATTGATTAATAATTAACATGAAATCTAAATTAATTAATTAATTAATTAATTAATTA 9516
Db 2498 AsnMetLeuProLysIleCysGluAsnVal-----LeuLysThrAlaGlnGluAspGlu 2516
QY 9517 TCTTCAGTATATATAGCAATTAACATCTGAAAGAAATGTTAGTCATATATTAATCA 9576
Db 2517 -----GlyGluCysValProAlaGluAsnSerGluProAla-----Ala 2530
QY 9577 AAATATCTCAATCGCTTGGAGTTAATGATTAATGAATAGTACAGAACAAA 9636
Db 2531 ThrAspSer----- 2533
QY 9637 AATTAATGAAAATTAATGAATTCAAAGATGTAATAAATAATTAATTCCTGTTATATTT 9696
Db 2534 GlySerGluThrProGluGlnThrProValLeuLys----- 2545
QY 9697 GTTGAAGATGAACACACAAAAATCATGACTAGATGGAATTAATAAAGAAAGACCA 9756
Db 2546 -----ProGluGluGluAla 2550
QY 9757 ACAGTGTGCTTAAGCACTATTTCTTACACCCATGATGATGTTTATCAACGA 9816
Db 2551 ValProGluProPro-----ProProProProGluGluAlaProAla 2565
QY 9817 CCTTATTCACACATCAGTACGACAAATATATCTTAATAAT----- 9861
Db 2566 Pro-----IleProGlnProGlnProProProProGlnLeu 2579
QY 9862 -----GATATATGAAAAGTATGCTGTGTTATGATGAGCGGTATAGT 9909
Db 2580 LeuAspAsnProHisValIleuThrAlaLeuValIleThrSerThrLeuAlaTrpSerValGly 2599
QY 9910 TTG-----ATAGCGCTTCATTTTCATGAAGAAATTAATCAATCTCTGTG--- 9954
Db 2600 IleGlyPheAlaThrPheThrTyrPheThrLeuLysLysLysThrLysSerSerValGly 2619
QY 9955 GACTGTGTGCGTATACGATATATCCCGCAAGAGAGATGGAATGCCATGCTGGAATCC 10014
Db 2620 AsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThrLysLeuSer 2639
QY 10015 AAAAATAGTATCATCATATATAGAGTGTCCATATTAAGCAAAATATATATATG 10074
Db 2640 ProAsnArgTyrIleProTyrThrSerLysTyrArgGlyLysArgTyrIleTyrLeu 2659
QY 10075 GAAGAGATGAT-----AGTGAAGATGAAGTAAATATATGTTGGAATTTATCTCC 10125
Db 2660 GluLysAspSerGlyThrAspSerGlyTyrThrAspHisTyr----- 2673
QY 10126 TCTGATATTAATTCATCGAAGAGTATGATGAAGAAATGATTAATGATATATATGTA 10185
Db 2674 SerAspIleThrSerSerGluSerGluTyrGluGluMetAspIleAsnAspIleTyrVal 2693
QY 10186 CCAGTATGCTTAATATATAAATATGATGATGATGATGATGATGATGATGATGATGAT 10245
Db 2694 ProGlySerProLysTyrLysThrLeuIleGluValIleGluProSerGlyAsnAsn 2713
QY 10246 ATACCAAGTAT----- 10257

[illegible][illegible]

QY	3220	GAGAAGCTGTAGAGCTGTAGAGATAGAGATATAGCTCAAGCGCTGTACGAAAGAGAGTGT	3279
Db	1060		1079
QY	3280	ACAGCTGTACGAGATGACACAGAGCTGTATGATGATTAATGATTAATGATTAATG	3339
Db	1080	Val---CysLysProCysAlaIalysGlnCysLysGluTrpLysGlnLysIleLysAsnTrp	1098
QY	3340	AAAGACAAATGGAAT	3399
Db	1099	GluGlnIleIleTrpArgLysIleSerAspLysIleLeuLeuTrpLysGlnLysIleLysAsnTrp	1118
QY	3400	TCTGTAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAG	3453
Db	1119	ThrAlaArgAsnIalGly-----AlaThrSerPheAspGlyThrAspProAsp	1134
QY	3454	---AGCAATGTATATGAAATTTTGTGCGAATTA-----TACCAACAAATAGTGTGC	3501
Db	1135	TyrLysGlnIalValAlaAspPhePheLysGlnLeuGlnIleLysSerIleLysSerIleLys	1154
QY	3502	AAAGTAAATTAAGTGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT	3561
Db	1155	LysArgProLysArgSerThrAspGlyThrThrAspProThrLeuThrSerProTrp	1174
QY	3562	GAAATGTGTGACATATCTCATATGATATACAGAAATTTGATGATGTATGATGATGATGAT	3621
Db	1175	PheThrAlaIalGlyTrpIleHisGlnGlnLeuProAsnMetGlyCysIleSerGlnThr	1194
QY	3622	GAGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3678
Db	1195	HisPheCys-----GlyAspAsnAsnLysLysGlnTrpAlaPheArgAsp	1209
QY	3679	AAACACAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3738
Db	1210	LysProThrAspHisAspLysValCysAspCysGlnSerIleGlnLys-----Leu	1226
QY	3739	CAGATTAATTAAGTGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT	3798
Db	1227	GlnTrpLys---LysLysGlnGlyIleValIleGlnGlnProCysAsnValIalLysThr	1245
QY	3799	ATATCTTAAGTGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTATAGTAT	3855
Db	1246	LeuLeuLysAspLysAsnGlnLysAsnSerAspIleGlyLysCysIleArgLysTrpLysAsp	1265
QY	3856	ACTAAT---GGATATCCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTCGATGTC	3909
Db	1266	GlyAsnGlnLysTrpValIleTrpAspCysThrArgAsnLysIleLysIleGlyGlnGln	1285
QY	3910	CGTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3969
Db	1286	GlyAlaIalTrpMetProProLysArgLysLysLysLysValAspPheLeu-----	1301
QY	3970	GAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4023
Db	1302	-----LysGlnLeuLysAspLysIleThrAspGlnLysLysLeuArgAspAlaPheIleLysSer	1319
QY	4024	GCAGACAGACGAAACATTTCTTCTCATGATGATGATGATGATGATGATGATGATGATGAT	4083
Db	1320	AlaAlaIalGlnTrpPheLeuSerTrpHisLysTrpLysGlnLysLysLysLysLysLys	1339
QY	4084	-----GAACTGCAATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4113
Db	1340	ProThrGlnGlnAlaProSerLeuAspValGlnIalGlnThrGlnLeuAsnAspGlyIle	1359
QY	4114	ATTCCTCCCGCATTTTGTAGATCATGTCATACATTTTGTAGATCATTTTGTAGATCATTTT	4173
Db	1360	IleProGlnGlnLysPheLysArgGlnMetPheTrpPheGlyAspTrpArgAspLys	1379
QY	4174	TTTGAACAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT	4233
Db	1380	LeuGlyLysAspIleGly-----AsnAspValProIleValAsnAsnIleThr	1396

Db 1507 ----- 1507
QY 5356 ACTACAGAAATCCCGTAGTACTGCGCGAAAATTTTCTGGAAGCAAAATAGAAATGT 5415
Db 1507 ----- 1507
QY 5416 GGTGGAACGCAATGATATGCGGGTACAAACCGTGTAGGATGATGAAATAGTGAAT 5475
Db 1507 ----- 1507
QY 5476 AGTGAAGAAGTGAATGATCTAATAAATGCTTCTGTACTCTTCAATGATGATAT 5535
Db 1507 ----- 1507
QY 5536 CCTATGGGAAAAATCGCATGAGTACTGCGTATCACTTCTGATGTTGCCGAA 5595
Db 1507 ----- 1507
QY 5596 TGGGGTGAAGATTTTTCAAACATAAAGAAAGAAATTTGGAGAAATTTGGAGCGCTGT 5655
Db 1507 ----- 1507
QY 5656 AATGATTACTTGTGTGATATGAAGATAAAGAAATGTACAGATGCGGTACA 5715
Db 1508 AsnAspGlySerCysHisLysAsnGlyGluGluCysGlnLysCysArgGluGluCysLys 1527
QY 5716 CAATATAAAAATTTTATAGTGTGAGAAACGACATATGAAACAAATCAAAAATAT 5775
Db 1528 LysTyrGlnLysTyrLeuGlnLysTyrPargLysAsnTyrAspLysGlnLysValLysPhe 1547
QY 5776 GGTGAAGATAAAGACAAAATATATTCGAGCATCTCTGTGCAAAATATGACAGACGCT 5835
Db 1548 LysThrAspLysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1566
QY 5836 CGCGAATATTTTATGACAAACATTAATAAATAAT-----TGTGAATAATAAAGTGA 5886
Db 1567 TyrGluTyrLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1586
QY 5887 GATTGTGAATATATAGTATGAAAGATGTGTCCACACAGCATTAACGTAGTAAAT 5943
Db 1587 AspCys-----LysCysMetGlnGluLysSerLysGlnGluSerHisAspLysSerIle 1604
QY 5944 AGTCAAAATATATGCCCCCATCATATAGACATGAAACCAAAAGATTTGAAGAAATGTAA 6003
Db 1605 ThrAspAsnMetProGlnLysLeuAspLysGluProGlnGluValArgGlyLysCysThr 1624
QY 6004 TGTCAAGTGCACAGAGTGCACACGTCGTACGAAGGAAACACGCTCACACAGCGGTATCA 6063
Db 1625 CysAspSerProThrLysLysProGlnValLeuProValSerPro----- 1639
QY 6064 CTGATATCAAAAGCGAGCGCATGAAAAAAGAAAGCAAGAAACAGCGCGCTTACAAAACAG 6123
Db 1639 ----- 1639
QY 6124 CCGAAAAAAGTGAATAATCTAACACAGAAATCGGACCAACACAGAACCCGACGACGA 6183
Db 1639 ----- 1639
QY 6184 GCACAAACAACAGAAAGCAACATCAACAGACAAACAACAGAAATCTGACGTGGCACA 6243
Db 1640 ----- 1647
QY 6244 ATGGTAAAGCCATTCTTTCGAAATAAACAGATAGCAGGGGTGAATAGAGGGTTGTAAT 6303
Db 1648 IleValAspIleLeuAsp-----ProAsnSerGlyThrGlyLysAsnAspAlaCys 1665
QY 6304 CCAAAACGATGGACAA-----TATCCTAAATGGGGTGTATGTATAGAGTAAG 6351
Db 1666 GluGluLysTyrGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 1685
QY 6352 TCTAAA-----GAAATGAAATGCAATATGTATGCTCTCTAGGAAAAAATATATGT 6405
Db 6405 ----- 6405

Db 1686 IleGluLysGlyLysGluAspAspValValCysIleProProArgArgGlnLysLeuTyr 1705
QY 6406 ATAAATTAATATACAAATATTAATTAATGAACCTGAAAAATAGCGTACATGATATATAA 6465
Db 1706 IleLysGlnLeuLys-----GluPheSerGlyLysLysAsnAspGluLeuArg 1721
QY 6466 GAGCGTTTATTAATAGTGCAGCATATAGAAACCTCAATTTTGTGCGTTAAATATATAT 6525
Db 1722 LysAlaPheIleGluCysAlaIleValGluThrPhePheAlaThrPheLysTyrLysGlu 1741
QY 6526 GAA----- 6528
Db 1742 AspLysLysAsnGlnGluLysLysGluGluThrSerTyrLeuValArgLeuGlyPheLeu 1761
QY 6529 -----AATCGTGCAGAAAAATGCAAAATGCAAAATGCAAAATG 6567
Db 1762 GlyLysGlnLysSerProAspGlnGluGlnLysAspLeuGlnLysSerGlyLysIle 1781
QY 6568 CCGATGATTTTAAAGATATGATATATATATATATATATATATATATATATATATATAT 6627
Db 1782 ProAspGluPheLysArgLysMetPheTyrThrPheGlyAspTyrArgAspIlePhePhe 1801
QY 6628 GAACTGATATTTCTAATGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 6687
Db 1802 GlyLysAsnMetCylLysAspMetLys-----ThrValGluLysLysLysSerIle 1819
QY 6688 CTCATGAAAAATATATAAAGAAAAACAGATATATAAATAAATAAATAAATAAATAAATA 6747
Db 1820 Phe-----ProAsnSerValLysProAspGlnLysThrArgGluThr 1833
QY 6748 TTTTGGAGAAAAATTAATAAATTTTATTTGGAGAAATGATATATATGATTAATCTTATCAT 6807
Db 1834 TrpTyrAsnAsnAsnAlaLysAspIleTyrPasnGlyMetLeuCysAlaLeuSerTyr 1852
QY 6808 CTCACAGCAAGAAAAAAGAAAAAATAAGATATATATAC-----CAGTAC 6855
Db 1853 -----AspThrLysLysAspArgLysLysLysLysLysLysLysLysLysLysLys 1870
QY 6856 AATGACATGACCAACTGACG-----CCTTCCCTGGAAGATTGTAAAA 6900
Db 1871 AsnAspTyrAsnLysValThrPheAspAsnGlyAsnThrLeuGlnLysIleTyrAlaThr 1890
QY 6901 AGCCCCAATTTTGAATGCTGTACACAGATGGGCGACAGAAATTTTGTATTAAGAGAAC 6960
Db 1891 ValProGlnPheIleLeuSerPheGluGluThrAlaGluLysPheCysArgLysValGly 1910
QY 6961 GAACAGTTGTTAAATTTGAGCGCGGCTGTAG-----GAAATATGAGTGT 7005
Db 1911 HisLysLeuAlaHisIleLysAspAspCysArgSerAspThrProGlyLysMetTyrCys 1930
QY 7006 AATGTATGATATGACGGTAAAGACACAGAA----- 7035
Db 1931 SerGlyAspGlyHisAspCysThrAsnGlnGluLysHisAsnAspMetPheAlaAsp 1950
QY 7036 -----TGTGACAGCGGCTGTAAATATCAATATCAAAATTTTATTAAGAACTG 7080
Db 1951 SerTyrCysProAspCysLysLysAlaCysArgLysTyrAsnLysTyrIleGluLysLys 1970
QY 7081 AAATCAATATGAAAGCAAGCAAGAAAGAAAGTCAAAAGATTA----- 7125
Db 1971 ValGluGlnPheTyrAsnGlnLysSerLysLysLysLysLysLysLysLysLysLysLys 1990
QY 7126 -----GATGCAAAAAGTATAGATATCTTCTACAGAAAGACATAGAGAAAGCA 7179
Db 1991 AsnLeuAspAsnLysTyrValLysGluPheTyrAlaThrSerGlnGly-----LysTyr 2008
QY 7180 ACATGTCCTCATGATATTAATTAACATGAATTAAGAAATTAATGTGCAATATAGATGT 7239
Db 2009 LysSerValAspSerPheLeuAsp-----LysLeuLysGlu----- 2020
QY 7240 TCTTGTATGCAAAAACCTTCTTCAACATACCAAAAAACAACACATCATCATCTCC 7299
Db 2020 ----- 2020

QY 7300 GATGCTATGATATGCCAGATCCGTGATATGTCCTGAGAAATTTAAACAGCTGAC 7359
Db 2021 -----ArgSerHis 2023
QY 7360 TGTCCGTAACCTTTCAAAAAAGGATCTATGATTCATCAAAAAAATTACTGAACCTTAA 7419
Db 2024 Cys----- 2024
QY 7420 ATACCTATGATGTGTAGAGAAAGCAGCATATTTATTTATCTAAGAGAGAAATTAAT 7479
Db 2025 -----His 2025
QY 7480 ATGATATTTACCTTGAAGAAAAATTTATACCTATGAGTCTCAAGAGAAAGAAAGT 7539
Db 2026 MetAsp---ThrLeuGluGlyLysIle-----AspPhe 2035
QY 7540 AAAATGCTGGAGCATATATATCCCTGGATCCCTAAGAACCTTATGCACTGATATA 7599
Db 2036 LysAsn-----ProLeuLysThrPheSerSerSerThr 2046
QY 7600 TATATAGAGAGAGAAACCT-----TGTGAATAATGAGAGAAATCGT 7644
Db 2047 TyrCys---LysThrCysProLeuTyrGlyValGlnCysArgAsnThrSerSphis--- 2064
QY 7645 TTTAAGTATGATATGATGAATGGAATGTTACAAAAATTCAAAAGTCTATGAGAGAAAAA 7704
Db 2065 -----CysIleGlnAsnSergly-----AsnGluLysLys 2074
QY 7705 AGATATGTGTACCTCCAGAGAGAAACATATGTCTTAAAGAAATTTAAGTAATTA 7764
Db 2075 -----TrpGluHisAla-----LeuAspThrIleLys 2083
QY 7765 ATGGAAGACTTAAAGATATGATATATTCCTCAAAAAATGTTGTCGACAGTCAAGCAAT 7824
Db 2084 IleLysAsnGlyAlaProThrSerIleAsnValGlnMetIleAspArg----- 2099
QY 7825 GAAGATAGACATATATAAAAACTTCACTGAGAGAGAGGGTGGCAATGAATCAATA 7884
Db 2099 ----- 2099
QY 7885 TGTGATACTATGAATATATGTTTCGCTGATGCGGTGACATAGTAGAGAGACAGATATG 7944
Db 2100 -----ArgGly----- 2101
QY 7945 TTACGAATTTGTGTACTTACTCCCGTAGAAATTAATTTATTAAGTTTGAATAC 8004
Db 2102 -----GlnTyr 2103
QY 8005 ATATATGAAAAATGAGAAAAATTAATAAGGTAGAAATTAATACAGAGATGACAAAG 8064
Db 2104 Ile-----GlnGluHisSergLysAsnSergPheLysGluSeraArgLeu 2117
QY 8065 TTTCCTTCTGCTGTGGGATGCTATATAGAAAAGATATTTGG----- 8106
Db 2118 LeuLysSerVal-----ArgGluGlnLysTrpGluCysSerPheValAsn 2132
QY 8107 ---AAAGCAATGACGTGCAAAAGCAGACAGAAAGTGAACCTTTAGAAAAGAAAGATG 8163
Db 2133 LysLysMetAspValCysLys-----LeuLysAsnAspPhe---LysGluAsnIle 2147
QY 8164 GATGATTTGAACGCATACAC-----TTAATACAGATAGATGCGACATTAAGAC 8214
Db 2148 AspThrAspGluTrpIleThrPheLysValLeuLeuGluAsn----- 2161
QY 8215 GATCCACCTGTGATGATATATATACCTCAACGCTTTGATGATGACTGAATGGTCTGAA 8274
Db 2162 -----TrpLeuGlnAspPheIleGlu 2168
QY 8275 ---TATATTTGTAAGACAGTGAAGAAATTTGAAAAATTTAAATATATGATGACAC 8331
Db 2169 GlyTyrTyrIleSer-----LysArgLysIleAspIle 2179

QY 8332 TGTAAAAATCTGCAGACAGATGCAAGAAATGATATGATGAATAATAGTGAACAGTATAA 8391
Db 2180 CysThrLysLysGlu-----GlnHisThrAlaIleGlnLysLys 2193
QY 8392 ACGAGATGTCAGAAATATATAAATTTTGTCTTAATGGAATCTCTATGATATACAA 8451
Db 2194 SerLysCysGlu-----CysIleGlyLysTrp----- 2202
QY 8452 TCMAATTAATACAAAGAAATGTATGACAAACCAATATACAAAATCTCTACTATGAT 8511
Db 2203 -----LeuLysGlnLysThrThrGluTrpAsp 2211
QY 8512 CAGTTCAAAATTTTGTCACAAAAGTTGAAAACCTTTTAAAGTGAATGTCTGTGAGAGC 8571
Db 2212 GluIleLysThrHisPheAsnLysGlu----- 2220
QY 8572 TTTTGTGATATCTTCATGAAACAGTAAGTGTGATTAATTAATGAATAATGAT 8631
Db 2221 -----AsnArgGlyAsp 2224
QY 8632 GGTTCCTCCATATATAGAACATATGCTTTCGAGAAACCCAAAAGTTATTAAGAGCT 8691
Db 2225 Gly-----TyrGluIleAlaHisLysValArgAsnTyrPheGlu--- 2237
QY 8692 TGCAGTTGACACTACCTTCAAGAAATCCATTTGGATATTTGCTACCGATCAAAACAAA 8751
Db 2238 -----LysAsnAlaVal----- 2241
QY 8752 GATGATGTAAAGAAATTAACAACTTTTACCTTGTCTGAGAAATGATATGATATAAT 8811
Db 2242 -----Gln 2242
QY 8812 CTGATATTTGAAGCGCATACCTTGTCTTATATGTTCAAGTGAATTAACAAGGTATG 8871
Db 2243 LeuLysLysTrp-----IleAspPheLysLysHisVal--- 2253
QY 8872 ATTCCTCCAAAGAGACATTTATGTACAAAGCTATGCTGATGCAATTAATTAAGAAA 8931
Db 2254 -----LysLysIleAspAspLeuArgIleValLeuTyrAspCysAlaAsn 2269
QY 8932 GGTGATTAAGAAATTTTAAAAAAACTTCTTACTTCTGCTTCAAGACAGCAATG 8991
Db 2270 MetAspLysGluIleLysProPheGluAsnGlnProHisGluThrSerProAsnTyr 2305
QY 8992 TTAGTCAAAAATTAATGGAAGAGAGATGCTTGAAGCAATGGAATATAGTAT 9051
Db 2286 LeuLysLysGluIleLysProPheGluAsnGlnProHisGluThrSerProAsnTyr 2305
QY 9052 GCAGATTTATCCGATATATTAAGAGAGAGATGATGAGACACTTCATTATCTGAAAAA 9111
Db 2306 CysAspIleSerProThrHisIleLeuThrGluProSergLysAspThrGluTrp 2325
QY 9112 ATTAATAAATAATTTTGAACATCAAAATGAGCAACGAAAAATGTAACATGCTGGAA 9171
Db 2326 LeuAsnProTyrAspGluTrp----- 2332
QY 9172 AATTAATAGCGTAGATATGCGACGCTATGTGTGATATTAATTTGCTACTTCAAA 9231
Db 2332 ----- 2332
QY 9232 GTAAACATTAAGATGAAGATGCTCAATTAACAAAGATGAAGAACTAATGATTTCTT 9291
Db 2333 -----ProGluAspAspThrSer----- 2338
QY 9292 CGTTGTTAATGAATGGCAAGCAAGCATGTAAAGAAAAAACAATTAAGTATTA 9351
Db 2339 -----ThrSerSer 2341
QY 9352 TTAATAACAAATTTGCTCTGTTCAACGAAGATTAATTTGAAGCGTCAGAAATTAATA 9411
Db 2342 ArgProAsnPheCysPro----- 2347
QY 9412 CAACCTGATGTCAGAAATGATATTAAGAAATATATTTAGCTTGAATATATGATAAAAAT 9471

```

Db 2347 ----- 2347
QY 9472 ACNATGAAATCTAATATATAATAGCAATTAAGATCAATCTCAGTAATATA 9531
Db 2348 -----GlnValGluProProProLysProGluVal 2357
QY 9532 GACATTAACATCTGAGAGAAAATGTCAGTCATATATAATCAAAAGATTCATAGC 9591
Db 2358 ProGlnLysPro---GluGluThrAlaGluAspThrThrGluAspThrGluGluVal 2376
QY 9592 GCTTGAGATTAAATATATATATGAAATAGTACGAGAACAAAATATGAAATAT 9651
Db 2377 AlaAlaPro-----ProValAlaProSerSerGluGluGlu 2389
QY 9652 GAATTCAGAGAGTACTAAAAAATATATCCGTGTTATATTTGTGAAGATGAACA 9711
Db 2390 AlaProLysGluValValProGluLysLysPro----- 2400
QY 9712 CACAAAATCATGTACTAATGAAATATAAAGAAAGAACAGCAAGTCTCTTAA 9771
Db 2401 -----LysGluValProLysProGluProLys 2409
QY 9772 GCACCTATTTCTTACACCCATGATGATCTTCTATCAAGCAGCTTATCTCAACA 9831
Db 2410 Ala-----ProLysLysArgArgProArgGluValThr 2420
QY 9832 CATCGAGTAGCAATATGATCTTAAATAATGATATATGAAAAGTATCTCTGTT--- 9888
Db 2421 His-----SerIleLeuProGluMetValSerIleSer 2431
QY 9889 -----GTTATGTATGCGCGCTTAGGTTGATAGCGCTTCAATTCATGAAGAA 9936
Db 2432 AlaPheProLeuSerValGlyIleAlaPheAlaLeuSerTyrPheValLeuLys 2451
QY 9937 AATTAATGATCGTGTGCTGCTGCTGATATCGAATATCCGGAAGAGATGGA 9996
Db 2452 LysThrLysSerThrIleAspLeuArgValIleAsnIleProLysGluAspTyrGly 2471
QY 9997 ATGCCATCTGCTGATCAATCAAAATAGTACATACATATAGAGATGCTCATTAAGGC 10056
Db 2472 IleProThrMetLysSerLysAsnArgTyrIleProTyrAlaSerAspArgTyrLysGly 2491
QY 10057 AAACATATATATATGAGAGAGATCTAGTACGATGAGATGAATATATATGAGGAC 10116
Db 2492 LysThrTyrIleTyrMetGluGluAspGluSer---AspAspTyrThrTyrIleGluAsp 2510
10117 TATCTCTCTGATATCTACT---TCATCCGAAAGTGAAGATGAAGATTG 10164
2511 IleserSerAspIleThrSerSerGluSerGluTyrGluMet 2527
RESULT 9
Q25733 PRELIMINARY: PRT: 2924 AA.
AC 025733:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE PfEMP1 variant 1 of strain MC.
GN MCVAR-1 PfEMP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MALAYAN CAMP;
RX MEDLINE=95530812; PubMed=7541722;
RA Baruch D.I., Pasloske B.L., Singh H.B., Bi X., Ma X.C., Feldman M.,
Tarsach T.F., Howard R.J.;
RA "Cloning the P. falciparum gene encoding PfEMP1, a malarial variant
antigen and adherence receptor on the surface of parasitized human
erythrocytes."
RL Cell 82:77-87(1995).

```

```

DR EMBL: U27338; AAB60251.1; -.
DR InterPro: IPR000345; Cytic_heme_bind.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP: 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
FT VARIANT 104 104 S->T.
FT VARIANT 105 105 E->K.
FT VARIANT 139 139 D->G.
FT VARIANT 362 362 K->E.
FT VARIANT 503 503 K->N.
FT VARIANT 581 581 S->T.
FT VARIANT 593 593 M->V.
FT VARIANT 617 617 I->K.
FT VARIANT 631 631 V->G.
FT VARIANT 658 658 D->E.
FT VARIANT 722 722 K->N.
SQ SEQUENCE 2924 AA; 335860 MW; 7077553BDAC82B26 CRC64;

Alignment Scores:
Pred. No.: 3,95e-151 Length: 2924
Score: 3154.00 Matches: 1013
Percent Similarity: 38.74% Conservative: 453
Best Local Similarity: 26.77% Mismatches: 963
Query Match: 16.21% Indels: 1356
DB: Gaps: 130

US-10-087-013-1 (1-10628) x Q25733 (1-2924)
QY 97 AAAGTGAAGAAATGTTTGGACGTATGCCAAAATATA----- 138
Db 13 LysAspAlaLysHisAlaLeuAspArgIleGlyGluValTyrLysGluValGlu 32
QY 139 AGACATCATCAATCAAAATATGCAAAAGAAATGATGATGCTGTTGAAGAGGATTTGACGAAA 198
Db 33 AsnAspAlaGluLysTyrLysLys-----AlaLeuLysGluLysLeuGluGlu 48
QY 199 GCAGAAATTCGTGCTGCTCTTACGCCAGTAATATAGCATATATATATATATATATATAT 258
Db 49 AlaLysGlyIleGlyLeuAlaLeuAlaSer-----ProAsn 60
QY 259 CCATGTAATTTAGATCATAGAAATCAATATAT---TTACGGTATGATGATGATTTG 315
Db 61 ProCysLysLeuValGluAspTyrTyrAsnAspArgLeuLys-----ArgLys 76
QY 316 AGACATCTCTGCTGATGAGAGAAACAAACGATTTGATGATGAGATGATGATGATGATGAT 375
Db 77 ArgTyrProCysAlaAsnArgIleThrValArgPheSerAspLysTyrGlyGluGluCys 96
QY 376 GGA---ATTAATATGATATTTAAAGAAATATGATCTATA---GCCTGTGGCCCA 429
Db 97 ThrPheAsnArgIleLysAspSerGluAsnAspAspSerIleGlyAlaCysAlaPro 116
QY 430 CCTAGAGAGACATATATGTTGATTAATAAACTGGAAGCTTAAATGATATATATATATATAT 489
Db 117 TyrArgArgLeuHisLeuLysAspTyrAsnLeuGluLysMetLysThrSerThr 136
QY 490 AATATTCATGATTTTGGAAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 549
Db 137 Lys---HisAspLeuLeuAspValCysMetAlaAlaLysTyrGluGluLysAspSerIle 155
QY 550 GTTAATATATAT---CCACATAAGAACAT-----TCAGACGCTTGATCT 551
Db 156 LysThrHisTyrThrLysHisGluLeuThrAsnProAspThrLysSerGluLysCysThr 175
QY 592 GCTCTGACAGAGTTTTCAGATATAGAGATGATGATGATGATGATGATGATGATGATGATGAT 648
Db 176 IleLeuAlaLysSerPheAlaAspIleGlyAspIleAlaArgLysAspLeuTyrLeu 195
QY 649 -----AAACCAATGTCCATGACGAAAGATGAAGATGACGATGATGATGATGATGAT 693
Db 196 GlyTyrAspAspLysGluLysAspGluArgLysLysLeuLysLeuLysLeuLysLeuLys 215
QY 694 TTCAGAAAATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753

```

Db 216 PheylsLysIleHisGluAsnLeuGlyThrGlnAspAlaLysAspHisLysLysAsp 235
QY 754 TCTGGAATTTTAAATTAAGAGCATGTCGCAATGGAATTAAGATTAAGTTCG 813
Db 236 GlnLysAsnLysLysGlnLeuAlaGlnAspTrpThrAlaAsnAspSerThrValTrp 255
QY 814 GAACCTATTAACATGTGATGCATATTAATCTGATATTTTATGCAATAGAAAGTAT 873
Db 256 LysAlaIleThrCysHisAlaGlnGluSerAspLysLysThrGlnGlySerCysSer 275
QY 874 ACACCATTTATTTCAATCCCT--AAATGCCCATTAACAGAGAA-- 918
Db 276 GlyLysLysLysAspLysCysAlaGlyCysLysAspGlnLysLysAsnGlnThrAsn 295
QY 919 --GTTCCCTACCAATTTAGATTTGTCCTCAATATTTAGCTTTGTTGCGAGTTCGGA 975
Db 296 GlnValProThrLysPheAspLysValProGlnLysLysLysLysLysLysLysLys 315
QY 976 GAAGAGTTTGGCCGAAAGAAAGAAATTAATTAAGTAAAGGTCAGAGCTGCTGTAAT 1035
Db 316 GlnAspPheCysAlaGlyLysLysLysLysLysLysLysLysLysLysLysLysLys 335
QY 1036 GACAAA--GACGCTATATTTGATGATGATGATGATGATGATGATGATGATGATGAT 1089
Db 336 GlnLysGlyLysAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 354
QY 1090 TGGAAAAAGGATTTTGGATTTGATTAATTAAGTCTAGCTGATGATGATGATGATGAT 1149
Db 355 ArgGlnAlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 374
QY 1150 GTTTTGAAGTTTGGTTAGGATCAACAGAGATTTAAAGAAAGAAAGAAAGAAAGTAT 1209
Db 375 ArgPheValLysTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLys 394
QY 1210 GAAAAAGAAATACA--TCATATTTATGCAAGATTAACAAATTTGCAAT 1257
Db 395 AspGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 414
QY 1258 AATATTAATAGTGAATATTAACAATTTTGAAGAAAGTAAAGAAAGTAAAGTAAAGTAA 1317
Db 415 AsnLeu--TyrValGlnHisPheLysLysLysLysLysLysLysLysLysLysLys 430
QY 1318 ACTAATGACACTTTTAAATTTACTAATTAAGAGAAAGTATTTGAAA--GGAAGA 1371
Db 431 ThrValAspLysSerLeuGlnLysLysLysLysLysLysLysLysLysLysLys 450
QY 1372 TTACAGAGGAAAGATATTTACTTTTACTAAGCTGCTGATGACAAAGGATATTTTAT 1431
Db 451 ValGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 470
QY 1432 CGTTGAGATATTTGCCAAGTGTCCGCACTGCGGCGTCAATGAT-- 1479
Db 471 HisThrThrLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1490
QY 1480 GGTATTAATATCACACAAATCA--GATATGATGCTGATGATGATGATGATGATGATGAT 1536
Db 491 GlnTrpLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 510
QY 1537 TATAAACCCTCAGTGGGTGTAAGCTCTAATATCACTGCTTTATAGGTATGAA 1586
Db 511 --AsnSerThrAspLysLysLysLysLysLysLysLysLysLysLysLysLysLys 523
QY 1597 CAAGCTGATTTACACAAATTAAGAAATTTTGTAAAGCTCACTAATTAATCAAGAT 1656
Db 524 ArgSerLysThrLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLys 542
QY 1657 AAAATATATCAAAATGGGATGCTATTAAGATGAAATATTAATTAAGATGATGATGAT 1716
Db 543 --AsnAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 555
QY 1717 GAACAAATATCTGAATCAATATGAT-- 1743
Db 556 AspAspGlnThrAspAspSerAsnAspCysValLeuGlnAspTrpGlnLysLeuThrLys 575
QY 1744 AATCCAGATATATATCATTTATTTTGAATTAAGGCTTACATTTATTAAG 1803
Db 576 GlnAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 595
QY 1804 GATACATTAATAGTGAATGACAAACTTAATTAATTTTATTAAT--ATATCAACACGAT 1860
Db 596 AspSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 615
QY 1861 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
Db 616 CysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 635
QY 1921 GAAGAAATGATATGATTAAGAACTGTTCAACAAAAAGAAATATATATATATATATAT 1980
Db 636 ThrGlnTrpGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 655
QY 1981 TAT 2040
Db 655 -- 655
QY 2041 AAGATGAGCAAAATGAGAAAGTATGGAATATTAAGAAAGAAAGAAAGTAT 2100
Db 655 -- 655
QY 2101 TCCAAATTTGAAATATATAGGCTATTTAGAGATGCAAT--GAACCTCTGTTA 2154
Db 656 --ThrHisAspAspPheLeuGlnLysLysLysLysLysLysLysLysLysLysLys 671
QY 2155 GATCACTTAAAGAAAGTGCACAGATATGTAAGCAATATATCAAGAAAGTATGTA 2214
Db 672 GlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 683
QY 2215 ACATCCCATATGCAACAAACACCCGTTGTTAAACCTGTCGAGCAGCAACCCACT 2274
Db 683 -- 683
QY 2275 AAAATATTAAGAAATATAGCAATATTTAAAGAGTGCATACAGAGCAACCAAT 2334
Db 684 --LysAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 694
QY 2335 CGTGGCTTCAATTAATGAAGAAAGCAACAGAGATATATATATATATATATATATAT 2394
Db 695 --GlnValLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 713
QY 2395 AGAAGGACTTCAAGGACATTTATGATGATATATATATATATATATATATATATATAT 2454
Db 714 AlaGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 729
QY 2455 GGTTCATTAATGACCAATGATGCAAGGACAGCTGATGATATACAAAGATTT 2514
Db 729 -- 729
QY 2515 GTGTAAGAACTGAATGGAAGTGAATCCGGAACATCGTAAAGATCAGAGATGTT 2574
Db 730 -- 730
QY 2575 ATTATGCTCTAGAGAGACATATATATATATATATATATATATATATATATATATAT 2634
Db 742 ThrCysProGln--Gln 747
QY 2635 GATCAACCACTTAATGATATATGTTGATGATATATATATATATATATATATATATAT 2694
Db 748 AspArgSerVal-- 751
QY 2695 GTTCTTATACGAAATATGAGCAAAAGATATATATATATATATATATATATATATAT 2754
Db 752 --AlaArgSerGlnSerAlaThrVal-- 759
QY 2755 AACCTAAAGGCCCCAAGAACTATGACCCCAACACAGACAACTATGTCGAGCT 2814
Db 760 --ProSerProAlaAspProLysAlaThrGlnLysLysLysLysLysLysLysLys 772

OY	2815	ATACGTTACAGTTTTCAGATATATAGTGAATATATATTCGAGGAAGAGATCTCGGAAAGA	2874
Db	772	-----	772
OY	2875	AACGGTACATGTGTAAGCTGCAAGACATTTGGAAACTGTTTTGGTAATATACATAMG	2934
Db	772	-----	772
OY	2935	TCACTCAAGAGCAAGGAATGATTAATATATATGATGATCCCCCAATTTAAATTG	2994
Db	772	-----	772
OY	2995	AGGGAATAATTGGTGGAGAGCTATAGAGCAAGATGTGGGAAGCATGAATGTGATATA	3054
Db	773	-----	777
OY	3055	AAATATTTGAAGATTAATTCGGACACCAATCACCAAAAGTAGTATTTGGGATATAGT	3114
Db	778	-----	778
OY	3115	GATCATTCACCATTTGGATGATTTATTCACCAAAAAATTAAAGTACATGACCGAATGGCA	3174
Db	779	-----	787
OY	3175	GAATGTACTGCAAGGTGCAAAAAAGAGTATAGTAACTGAAGAGAGAGTAAAGAG	3234
Db	788	-----	801
OY	3235	TGTAAAGATAAGATATATGGTCAAGGCTGTACGAA-----GAGAGTGTACAGCT	3285
Db	802	-----	821
OY	3286	TGTACGAAGTGCACAGAACTTTATGAAATATATGATATATATAGATTATAGAAACA	3345
Db	822	-----	824
OY	3346	CAATGATATATATATCAGATTAATACAAAGATTACATGAACAAACAAATGCTGT	3405
Db	825	-----	838
OY	3406	AGTAAATGTGATTATGAGCTTCACGACTGCCAAAAATCATATATGACAGAGATTAT	3465
Db	839	-----	847
OY	3466	GAATTTTTCGGAATATATACCAACAAATATGTCGCAAAAGATTAATAAGTGTCTAGT	3525
Db	848	-----	865
OY	3526	GATGAAGATGCTGTCATTGGTACTTAACACCAGTAAATAATGTTGAGCAATATCCAT	3585
Db	866	-----	880
OY	3586	GATACAGGAATTTTGATGATTTGTCACTCACAAATATGACTTTGTGATGAAAAAGTAT	3645
Db	880	-----	880
OY	3646	GGTAAAGATTAACGAAAAATATGCTTTAGAGATTAACACAGACCATGATGTCGCTGT	3705
Db	881	-----	886
OY	3706	GGTTGTAAGATGATGAAACGACAGAGGTACAGATTAACGAAAAAAACGCGAA	3765
Db	887	-----	891
OY	3766	GAAGAAGATACGGAATGTAAACAGTGAATGATATCTTAAAGAAAAAGATGGAAGAA	3825
Db	892	-----	894
OY	3826	CAAGTAGAAGATTGTCAATCCAAAAAAGAAATAGTAATGATATCCGATTTGCAATGCCGA	3885
Db	895	-----	895

[illegible]

```

Db 1247 CysHisIleGlnCysArgLysTyrArgLysTrpIleAspIleLysPheAspGluTyrHis 1266
QY 4618 CCACAAATATAATTTACAAAGATGACGACGACAAAAGATTCGATACACACATC 4677
Db 1267 LysIleGlnLysLysTyrGlnGlyIleTyrAspLysLeuThrLysAspLys 1283
QY 4678 GGTGTAATGCTTACAGACTACTGAGAACGAAATGCAACAGATTACTTGAACAGAAATTT 4737
Db 1283 ----- 1283
QY 4738 ACTGCTAGTGTGGTGATTAAGCCTGGAAGTCCCTGTGTGTACAAAGAAATTAACAAATTG 4797
Db 1284 -----SerSerIleGlyLysAsp----- 1288
QY 4798 TTGAAAAAACAGGCTTACTACTATGATGCCACAAACATTTGTGTGACAAAATTTATTGAA 4857
Db 1289 -----AsnAsnCysCysLysAspIleGlnLysHisLysSerAlaAlaValPheLeuLys 1306
QY 4858 AATGACGACAAATATATCTAATCACTTTCGAGTAAAGATTAAGTGAAGATTAGTAAAGAG 4917
Db 1307 Gln-----LeuLysHis 1310
QY 4918 GCAAACACAGCTGCTATTAGTGAACAAACAAAGCTTAACTACTACAACTTGA 4977
Db 1311 CysLysAsnGlyIleThrSer-----GluAsnLysGly-----AsnGlnLysPheLeuAsn 1328
QY 4978 GAATTACTGACATGCTGCTTTT-----TCCTTCCTGCTGCTACTGATATATGTTTCTATGC 5033
Db 1329 LysLeuAspPheAspLysIleProGlnThrPheSerProSerThrTyrCysLysLacys 1348
QY 5034 ATTGATGCAATTAATATACATCCAGATCCAGAACTTAAAGATGAATGGTTCGAAAAAGATT 5093
Db 1349 -----ProVal 1350
QY 5094 GATGCAAGTGGCGGACGACGAGAGGTACATTTGGTCAATCTACAAAAGAAAAAGA 5153
Db 1351 TyrGlyValAsnCysAsnGly-----AsnLysArg 1360
QY 5154 AAAAGAGAAATATAAACGCTGGATGCGACAAATATCTTATGAGTCCGCGCTTGTAG 5213
Db 1361 GlyArgGlyGlyThrAsn-----GlyCysThrThr----- 1370
QY 5214 TGCATGAATTAATGCTTTTATGATTTAAGATTAATTAATTCATGATTAATTTGGA 5273
Db 1371 -----AsnAsnGln-ProGlnAsnLysGln 1378
QY 5274 AGATGAAAAACAAAGACCGAGAGAAATTTGACAAAAATTTAAACAAAAATGGAATC 5333
Db 1378 uasnAsp----- 1380
QY 5334 AGTTGGCAAGAGATGATACTACAGAG-----AATCCCGTACTACTGCGCG 5384
Db 1381 -----LysGlyAlaAlaSerThrIleSerIleLeuIleAsnAspGlySerThr----- 1396
QY 5385 AAAATTTTTCGACAGCAAAATTAAGCAATGTGTGTGCAACCAATGATATCCGGGTACAA 5444
Db 1396 ----- 1396
QY 5445 ACGTGTAGGATGATGGAATATGGAATATAGTCAAGAAAGTATGAATCAATTAATAA 5504
Db 1397 -AsnGlyAlaIleThrAsnGlyThrThrGlyThr-----ThrAspGlnThrLeuLysGln 1413
QY 5505 ATGTGGTTCGTACTCTCAGATGATGATTAATCTATG-----GGGAAAAATCCGCGATGA 5558
Db 1413 uCys-----SerAspLysTyrAlaPhePheLysGlyLeuArgLysGlnGln 1428
QY 5559 AGTACTGCGTATGATTTCTTCGATGCTTTGCCGAATGGGAGAGCAATTTTTCGAAACA 5618
Db 1428 uTPTrpThrCysGlnLys-----LysTyrGlyValAsnGlnCys----- 1440
QY 5619 TAAAGAAAAGAAATTTGAGAAATTTGTAGGCGGTGTAAATGATTAATCACTGTGTGATTA 5678
Db 1441 -----AsnLeuThrAsnArgValAsnAsp-----ThrTyrPheAspLys 1453
QY 5679 TGAAGATTAAGAAAGAAATGTACAGATGCGTGCACAAATTAATAAATTTATATAGCA 5738
Db 1453 sAsp-----IleValPheAsnGlnPheGlnArg 1463
QY 5739 GTGGAAACACAGATATGAAAAAATCAAAAATATGCGTGAATTAAGCAAAATATATA 5798
Db 1463 gTrpLeuArgTyrPhe-----ValHisAspTyrAsnIleLeuLysHisLysIle- 1479
QY 5799 TTCGAGATCCTGTGGCAAAAGATGCAGAGAGACCTGCGAAATTTATGCAAAACAATT 5858
Db 1480 -----AspProCysIle 1483
QY 5859 AAAAAAATTTGTGAAATTAAGTGAAGATTGTAATATAGTATGTAAGAAAGATGTGC 5918
Db 1483 eLysLysGlnLysGlnAspLysThr-----GlnHisLysCysIle----- 1496
QY 5919 CACACAGGATTAATCTGATGTAATATGCAAAATATATGCCCCCATCATCATTAAGCATGAACC 5978
Db 1497 -----AsnGlyCysAsnIle----- 1501
QY 5979 AAAAGAGTGAAGAAAGTGAATTTGTCAAGTGCACAGAGTCCACACAGGTGACGAG 6038
Db 1502 -----LysCysGlnCys-----ValArg-- 1507
QY 6039 GGAACACCGCTCACACAGGATATCACTGATATCAAAAACGACGGCATGAAAAAGAAC 6098
Db 1507 ----- 1507
QY 6099 GAAAAACGCCGCCCTTACAAAACAGCCGAAAAAAGTGAAAAATCTAACACAGAAATCGC 6158
Db 1507 ----- 1507
QY 6159 AGCACAACACAGAACCCGACAGACGACACAAACAAACAGAAACATCAACAGACAC 6218
Db 1507 ----- 1507
QY 6219 AACAACAAATCTGACGTGGGACAAATGTTAAAGGCCATCTTTCGAATTAACCAAGTAG 6278
Db 1507 ----- 1507
QY 6279 CAGGGGTGCAATGAGGGTTGTAATCCAAAACGTATGCAATATATCCATATGCGGTTG 6338
Db 1508 -----LysTrpLeuGln 1511
QY 6339 TATTGTAGTAACTCTAAAGAAATGAAATGSCATATGATGCTCTTATGAGGAAAAA 6398
Db 1511 uIleLysGly-----AsnGlnTrpGly----- 1518
QY 6399 ATTATGATTAATTAATATCAATATTTAATTAATGAAACTGAAATTAAGCGTGACAAAGA 6458
Db 1519 -----AsnIleLys-----LysHisTyrAsnIleAsnSerHisAsn 1530
QY 6459 TATTAAGAGCGCTTTTATTAATATGTCACACATAGAAACTCAATTTTGTGTTAAATA 6518
Db 1530 pAspLysGlnThr----- 1534
QY 6519 TATTAATGAAAAATCTGCACAGAAAAATGAATTGCAAAATGGAACAAATTCAGATGAAT 6578
Db 1534 ----- 1534
QY 6579 TAAACAAATATATGATTAATATGATGATGATTAATAAGATATG-----TTTTTGAAGTGA 6635
Db 1535 -----IleAlaTyrAsnValLysSerTyrPheValAspGlnGlyLeuPheAspThrAs 1552
QY 6636 TATTCTATGATTAATAAATAATTAATTAACGTATACAAATATGTAACACCAATTCATCAATGA 6695
Db 1552 pTyrLysLysAlaGlnLysValVal----- 1560
QY 6696 AATATATTAAGAAAAACAGATTAATAAATAAAGATGAAGAAATTAACGTAATATTTTGGGA 6755
Db 1561 -----GluAspLysLysGlnLysLysIleThr-- 1570

```

QY 6756 GAAAAATAAAATTTATTTGGAGGAATGATATAGATTACTTATCATCTCAGAGA 6815
DB 1571 -----GlyCysThrGlyHis-----As 1576
QY 6816 CGAA---AACGAAAAAGAAAAATTTAGATATATTTCCAGTACATGACATGCCAACT 6872
DB 1576 pGlyCysSerGlyLysGlyLysGlyLysLys-----AspHei1leThrSni1e 1593
QY 6873 GACGCTTCCTTGAAGAGTTTGAATAAAGGCCCAATTTTGAGATGGTTCACAGAAATG 6932
DB 1593 u1leSerGlyLeuGlnAspLysIleThrSerCysGln----- 1605
QY 6933 GGCAGAGAATTTTGTAAATAGAGAGAACAGTGTAAATTTGAGAGCGGCTGTAA 6992
DB 1606 -----AsnLysHis----- 1608
QY 6993 GGAATATGAGTGTATGTATGATACCGTAAAGACACAAAGATGTCAGAGCGCTGTGT 7052
DB 1609 -----AsnProAsnGlyLysThr-----AlaCys-- 1616
QY 7053 AACATATCAAAATTTTATTATAGAGAGTGAAGAACTGATATGAAGACAAAGAGAAAGTT 7112
DB 1616 ----- 1616
QY 7113 CAAAAAGATAAGATGCAAAAAAGTATAGATTATCTTACTGAAAAGACATAGA 7172
DB 1616 ----- 1616
QY 7173 GAAGGCAACATGCTCATGATATTTAAACATGAAATTTAAAGATTTATGTGGCAATTA 7232
DB 1616 ----- 1616
QY 7233 GGATTTCTTTGTATGCAAAAAACCTTCTCACAACTACCAAAAAACACAAACATCACA 7292
DB 1617 -----AspProHeProSerProThrProGlnLuh 1627
QY 7293 ATCATCCGATGCTATGATATGATGCCAGAAATCGTGATATGTTCTGGAAGATTTAACA 7352
DB 1627 rAspProLeuAspAspAspThrProAspProLeuAspAspAspGlnHisThrGlnLuhInPr 1647
QY 7353 GGTGAGTGTCCCTGAACCTTCAAAAAAGGATCTATGATTCATCAAAAAAATTTACTGA 7412
DB 1647 oLysPheCysPro-----Pr 1652
QY 7413 ACCTAAATACCTATGATATGTTGTAGAGAAAGACGATATTTATTTCTAAAGACAGAGA 7472
DB 1652 oProProProProMetLThrCysValGlyLysIleAlaLysGlyLeuAlaGlyValaGlnaGln 1672
QY 7473 AATTAATATGATATTTACTCTGAAAGAAAA----- 7503
DB 1672 uGlyLysIleAsnAsnGlyLeuLysGlyLysAspPheAsnGlyLysCysAsnAs 1692
QY 7504 -----TTTATACCTATGAGTGTACAAAAG-----GAAAAGGAAAG 7538
DB 1692 nValLysLysLysAsnGlyAlaValaIleGlyLysGlnLysCysLysPheGlnLuhInThrTy 1712
QY 7539 TAAAAATAGTGTGACTATATATATATCTTCGATCCTTAAGAAACCTTATGCAACCTGATA 7598
DB 1712 rGlnAsnSerValaAsnAsnIleAsn----- 1720
QY 7599 AATATATAGAGAGAAAGCCCTTGTGAATAATAGAGAAAAATCGTTTAAAGTAGATTA 7658
DB 1721 -----AsnLysCysLysAspAsnGlnAsnGlnLuhArgPheLysIleGlyG 1735
QY 7659 TGAATGGAATGTTACAAATTCAAAGTCTTATCAGAGC-----AAAAAGAGATATGTG 7715
DB 1735 nLysStrp-----AsnPheLysStrpIleGlyThrIleGlyLysAspLeuLysI 1751
QY 7716 ACCTCCAGAGAGAGACATATGCTTAAAGAAATTTAGATGAATTTAAATTTAAAGACT 7775
DB 1751 eProProAlaGlyGlnHisMetCysLeuAspAspLeuSerMetLeuGlyArgThrThrI 1771

QY 7776 TAAAGTATGATATTTATCTCTTAAAAATGGTTCGTGCACTGCACAGAAATGAGGAATAGA 7835
DB 1771 eSerAspSerSerLeuLeuLysLysIleGlnGlnAlaLysSerGlnLuhArgAsp 1791
QY 7836 CATTAATAAAAAATCTTCACTCAGAGAGACGGGTCGCGAATGAAGATCCAAATTCATCTAT 7895
DB 1791 pIleIleArgLysLeuLeuGlnGlnAsnSerCysAspGlnHisArgIleCysAspAlaHe 1811
QY 7896 GAAATATAGTTCCTGATCTGATCTGGGATGACATAGTTAGAGACAAACATGTTGCAATTTG 7955
DB 1811 tLysTySerPheAlaAspLeuLysPleIleAlaGlyArgAspLeuThrAsnLysAs 1831
QY 7956 TGGTACTTACCTCCCGTAGAAATTAATATATAAGTTTGAATACATATATGGA 8015
DB 1831 nSerLysGlnLysGlyLeuGlnLysArgLysArgLysArgLysArgLysArgLysArgLys 1851
QY 8016 ATCGAGAAAT-----AAAT-----AAAGTAAATTAATTAACAGATGTACAAAC 8063
DB 1851 sLeuGlnAsnAspLysAsnLysStrpGlyLysAspArgProLysTySerLeuGlnLeu----- 1869
QY 8064 GTTTCGCTCTGCTGGTGGGATGCTAATAGAAAAGATATTTGGAAGAACTGACGTGCA 8123
DB 1870 -----ArgSerAspTrpTrpAspAlaAsnArgLysHisIleThrAsnAlaMetThrCysAs 1888
QY 8124 AGCACCAGAGATGCAAACTTTTAGAAAGAGAGATGAT-----GG 8168
DB 1888 nAlaProAspAspAlaLysPheLeuLysAsnProAsnAspThrSerGlySerSerSe 1908
QY 8169 ATTTGAACGCATATACATTTAATACAAATAGTGTGACATTAAGAGAGATCCACCTGTGA 8228
DB 1908 rSerLysGlyIleMetThrThrHisSerAsnCysGlyTyAspLysGlnProProAspTy 1928
QY 8229 TGATTTATATACCTCAACGGTTCGATGATGATGATGATGATGATGATGATGATGATGATG 8288
DB 1928 rAspTyThrLeuProGlnProPheArgTrpMetGlnGlnLysPheGlnLysPheCysLysLe 1948
QY 8289 ACTGATGAGAGAAATTTGAAAAATTTAAATAATCATGTGATGACTGT---AAACATCTGA 8345
DB 1948 uLeuAsnGlnLysMetGlnLysGlnLysThrCysGlyLysGlyLysLysAsnSerI 1968
QY 8346 CAGATGCAAGAAATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8405
DB 1968 eThrCysGlnAspAspArgAsnGlyThrAsnCysGlnAsnCysLysAsnGlnLysGlyL 1988
QY 8406 ATATTAATAATTTGTTCTTAAATGCAAAATCTATTCGATATGATCAATTAATTAACA 8465
DB 1988 sTyLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysL 2004
QY 8466 AGAATGTATGAAACAAATATAT-----ACAAAATCTTACTTATGATCA 8513
DB 2004 sGlnIleTyAsnGln---IleTyAsnAsnLysAspSerLysIleAsnSerAsnGlnL 2023
QY 8514 TGTTCAAAATTTTGTCAAAAGTTGAAACTTTAAAGTGAATG---TCGTGTGAGAG 8570
DB 2023 rPheLysLysPheLeuGlnLysLeu-----LysAspLysCysLysGlyLeuLysAsnSe 2040
QY 8571 CTTTTCGATATATCTTATGATGAAACAAAGTAAAGTGTGATTTAATTAATTAATTAATTA 8630
DB 2040 rSerAspLysCysIleAspGlnLuhThrHisCysThrLysTyLysPheSer---AsnSe 2059
QY 8631 TGGTCTTCCATATATACGAACATATGCTTTCGAAGAAACCAAAAGTTAATAGAGAGC 8690
DB 2059 rGlnAsnLysAsnHisAsnAsnTyArgPheLysAsnProProGlnLuhGlyLysAla 2079
QY 8691 TTGCACTGTACASCTCTTAAGAAATGATGATGATGATGATGATGATGATGATGATGATG 8745
DB 2079 aCysLysCysAspAlaPro-----AspProLeuAspAsnCysProLysAspSerAlaTh 2097
QY 8746 -AACAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8804
DB 2097 rTyGlyLysAlaCysAsnThrLeuLeuProThrLysLeuCysGlnLysLysThrPheAs 2117
QY 8805 TAATTAATCTTGATATTTGGAAGCAGATACCTGTGTTAATATGTTCAAGATGATTAACAAG 8864

```

Db 2117 nasnaspsaspsertpraptrhSerpheValGlnThrSerproArgaspsantrhG1 2137
QY 8865 TGTATTGATTCCTCCAGAGAGAGACATTTATGTACAGACCTTCACCTGCATATAATTA 8924
Db 2137 yValLeuValProProKarghArgGlnIleCysLeuLysasnIleThrThr---LysLe 2156
QY 8925 TAGAAAAGTGATAAA---GAAATTTTAAAAAAAACCTTACTCTGCTGCTTACGTCA 8981
Db 2156 uArgSerIleGlnLysIleAspSerPheLysAlaGlnLeuMetThrSerIleArgasnG1 2176
QY 8982 AGGCAATGTGTGGTCAAAAATTAATTAATGGAAGAAGACTTGTGCTTGAAGCAATGA 9041
Db 2176 uGlyLysLeuLeuCysGlnLeuLysLysAspArgAspValThrLeuGlnAlaMetL 2196
QY 9042 ATATAGTTATCCAGATTATTCGATATTAATTAAGAACTGATATGATGACCTTCAT 9101
Db 2196 sTySerPheLysArgLysArgLysArgLysArgLysArgLysArgLysArgLysArg 2216
QY 9102 ATCTGAAAAAATTAATAATTTT-----GAAACATCAAAATGAAAC 9143
Db 2216 oLeuAspLysLeuLysThrLysLeuAsnValLeuLysGlyAspGlyThrAsnGlnI 2236
QY 9144 AACCGAAATCTTAACATGCTGGGAGAAATATAGACCTGACATATGGACGCTATGT 9203
Db 2236 eLysGlnAspArgGlyLysArgLysArgLysArgLysArgLysArgLysArgLysArg 2256
QY 9204 ATGTGATATTAATTAATGCTTCAAAAGTAACTAGATGAGAGAGTGTCTCAATTAC 9263
Db 2256 uCysGlyLysLysAlaAlaGlyLysLysLysLysLysLysLysLysLysLysLysL 2274
QY 9264 AAAGATGAGAAACTTAACATGCTTCTGCTGCTTATTAATGAAAGGCAACCAAGATG 9323
Db 2274 o---AspAspAsnThrHisGlnPheLeuArgLysArgLysArgLysArgLysArg 2293
QY 9324 TAAGAAAAGAAACATGTAAGTATCATTAATAACAAATGCTCTGCTCAAGAGAA 9383
Db 2293 sAlaLysArgGlnLysLeuPheAsnGlnValLysArgLysLysLysLysLysLys 2311
QY 9384 TAATTTGAAGGCTGAGATTTATTAAGCAACCTGAGTGTGAG-----AATGATAT 9434
Db 2311 nCysIleIleGlnLysArgLysThrIleAspProValLysGlnLysAlaCysThrGln 2331
QY 9435 TAGAAAATTAATTAATGCTTCAAAAGTAACTAGATGAGAGAGTGTCTCAATTAC 9494
Db 2331 rArgAspLysIleThrArgLysIleGln-----G1 2341
QY 9495 ATATAGCAATTAATAAGATCATCTTCAGTAAATATAGCAATAAACCATTGAA---GA 9551
Db 2341 uTyrArgLeuLeuAsnLysArgLysArgLysArgLysArgLysArgLysArgLysArg 9552
QY 9552 AAATGTCAGTCAATATTAATAATCAAAAGATTCATGACGCTTGGAGTTAAATGATAT 9611
Db 2361 rLysAlaProGlnLysPheAsnAspLys-----CysAsnAspLysCysAsnCysLe 2378
QY 9612 AAATGAATTAATTAAGCAACAAAATTAATGAATAATTAATTAATTAATTAATTA 9671
Db 2378 uSerLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2389
QY 9672 AAAATTAATTCCTGTTATATTTTGTGAGATGAACACACAAAATCATGTACTAGA 9731
Db 2389 sAsnMetLysLysSerPhe-----AspAspAsnAspLysLysLysLysLysLys 2406
QY 9732 TGGAAATTAATAAGAGAGAGAAAGTGTCTCAAGACATCTATTTCTTATACAC 9791
Db 2406 sArgGlnIleLysProLysArgPro-----ProLysLysLysLysLysLysLys 2423
QY 9792 CCATGATGATCTTCTATCAAGACCTTATTTCTCAACACATGACATGACACAAATGA 9851
Db 2423 uHisThrProSerGlnGlnAspThrProProLysProLysProLysProLysProL 2443
QY 9852 TCCT-----AAAATGATATTAATAAGAGAGATCTCTCTGTGTAT 9893
Db 2443 oProProAlaGlnGlnProPheAsnArgAspIleLeuGlnLysThrIleProPheGlyI 2463
QY 9894 TGTATGCGGCTTAGTGTGATAGCGCTTCATTTGATGAAGAAAATTAATCAATGCTGT 9953
Db 2463 eAlaLeuAlaLeuGlnLysSerIleAlaPheLeuPheLeuLysLysLysLysLysL 2483
QY 9954 G---GACTTGTTCGCTATTAATTAATTAATTAATTAATTAATTAATTAATTA 10010
Db 2483 lLysAsnLeuPheGlnIleLeuIleIleIleIleIleIleIleIleIleIleI 2503
QY 10011 ATCCAAAAATAGCTATACATATAGATAGATAGATAGATAGATAGATAGATAGAT 10070
Db 2503 uSerProAsnArgLysIleProLysThrSerGlyLysArgLysArgLysArgLysArg 2523
QY 10071 TATGAGAGAGATTA-----AGTGAATGAGATGAATTAATTAATTAATTAAT 10121
Db 2523 rLeuGlnGlnLysSerGlyThrAspSerGlyThrAspLysLysLysLysLysLys 2538
QY 10122 TTCCTGATATTAAT---TCATCGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 10178
Db 2539 ---SerAspIleThrSerSerSerSerSerSerSerSerSerSerSerSerSer 2557
QY 10179 ATATGATCCAGTATGCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10238
Db 2557 eTyValProGlnLysSerProLysLysLysLysLysLysLysLysLysLysLysL 2577
QY 10239 AAGGATATTAACACTGAT-----GATACCAAGT-----AA 10271
Db 2577 yAsnAsnThrThrAlaSerGlyLysAsnThrProSerAspThrGlnAsnAspIleGln 2597
QY 10272 TGATACACACAGTGAATGATTAATTAATTAATTAATTAATTAATTAATTAATTA 10331
Db 2597 nAspGlyLysProSerSerLysIleThrAspAsnGlnLysLysLysLysLysLys 2617
QY 10332 TGTATCTCAATTA-----CCAAATACAGAAACAAATTAATTAATTAATTA 10382
Db 2617 eIleSerAsnMetLeuGlnAsnGlnProAsnThrGlnPro-----AsnMetLeuGly 2635
QY 10383 AGATATTCATGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10442
Db 2635 rAsnValAspAsnAsnThrHisProThrThr---SerArgHisAsnValGlnLysSer 2654
QY 10443 TTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10502
Db 2654 oPheIleMetSerIleHisAspArgAspLeuLysArgLysLysLysLysLysLys 2674
QY 10503 TAATATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10562
Db 2674 lAsnMet-----ValAsnAsnAspIleProIleSerAlaArgAsnGlyAsnTy 2691
QY 10563 AGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10622
Db 2691 rGlyIleAspLeuIleAsnAspSerLeu---AsnSerAsnLysValAspIleTy 2710
QY 10623 GATATTG 10628
Db 2711 LeuLeu 2712

```

```

RESULT 10
Q904A2
ID Q904A2 PRELIMINARY: PRT: 1685 AA.
AC Q904A2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 20, Last annotation update)
DE Variant surface protein pFEMP1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IT;

```


D	937	TrpAspLysAspGluGlySerLysLysMetAspValIleLeuLysLysIlePheGlyLys	956
Q	2926	ATCATATAGTCACATCGAAAGCGAAAGAAATGATAATATATGATGATGCCCAAAATAT	2981
D	957	IIeLysGlnGluIleuProLysGlnIleGlnLysLysTyrLysAsnProAspGlyLysHis	976
Q	2986	TTAAATATGGAGGAAATATGGTGGAGCGTAAATAGCGCAAAAGTATGGAGCCATGAAA	3041
D	977	ThnGlnLeuArgLysAspTrpTrpIleuLysAsnArgHisGlnValTrpArgAlaMetLys	996
Q	3046	TGCGATATTAATATATTGGAAGATTAATCGGAGACACATCAACACAAAGTAGTATTGC	3101
D	997	CysAlaIle-----GlnAspGlySerIleGluLysCys	1007
Q	3106	GGATATAGTATCATCAACCCATGGATGATATATCCACAAAAATTAGATGGATGACC	3161
D	1008	-----AsnGlyTleProLeuAspAspTrpIleProGlnArgLeuArgTrpMetThr	1022
Q	3166	GAATGGCGCAATGTGATCTGCACAGGTGCGAAAGAAAGAGATGATCAATGATTGAAGCGAAG	3222
D	1025	GluTrpAlaGluIleTrpPheCysArgLysMetGlnLysGlnAlaTyrHisGlnLeuLysGlyLys	1044
Q	3226	TGTAAAGCATGTAAAGATTAAGATTAATGTGTAAAGCGTGTACGAAAGAGGTGTACAGGT	3282
D	1045	CysSerGlnCysLysThrLysAsp-----LysLysCysThrAsnLysSer---AspAsp	1061
Q	3286	TGTACAAATGCGACAGCAAGCTGTGATGATGATATATATATATATATATAGATTTAGAAAGAA	3345
D	1062	CysAsnThrCysThrGlnAlaCysThrAlaTyrHisAsnArgLysIleAsnThrTrpLysGln	1081
Q	3346	CAATGGAATATATATCTCGATTAATTCAAAGAAATTACATGAAACAGCACAAATGTGTGTT	3405
D	1082	GlnTrpAspAlaIleSerAspLysTyrGlnPheLeuTyrLeuGlnAlaLysThrAlaIle	1101
Q	3406	AGTAAATAGCGATATGAAGCTTCCAGTACAGTCCGCAAAATCATATAGACAGGAATTATT	3465
D	1102	AlaAsnGlyGlyProHisAlaSerSerGlyAspValGlyLysAspLysProValVal	1121
Q	3466	GAATTTTGTGCGAATTAATTCACCAAAATAGTGGCAAA---AGTATTAAGAGTGGTCT	3522
D	1122	AsnPheLeuPheGlnLeuTyrLysGlnAsnGlyLysIleSerThrProSerAspThr	1141
Q	3523	AGTGATGAAAGCGCTGCATT-----GGTACTACACACACATGAA	3564
D	1142	HisProGlyProArgValLysArgGlyAlaProSerGlyAsnSerAspThrValTyrSer	1161
Q	3565	ATGTGTGGAGCATATCTCCATGATACAGCAAAATTGTGATGATGTGCAGTCAAAATATAG	3624
D	1162	ThrAlaIleArgLysTyrIleHisGlnGluAlaHisIleAspAspCysAsnLysGlnAsnVal	1181
Q	3625	TTTTTGATGAAAAAATGATGATGGTAAGAT--AACGAAAAATATGCGTTTATAGATATA	3681
D	1182	PheCysGlyLysLysGlyLysAlaAsnAspAsnAsnGlyLysTyrAlaPheHisProGlu	1201
Q	3682	CCACAGACACATGATGCGTGGCTGTGGTGAAGAAAGCGATGCAACGCAAGGCTACAG	3741
D	1202	ProTyrAspHisLysLysAlaCysAlaCysAspArgLysArgAsnProAspValLysVal---	1220
Q	3742	ATTAATAACGAAAAAAGCGGAAGAAAGATACGGAATGTTAAACAGTAAATGATATA	3801
D	1221	-----LeuGlnAspProCysLysMetValGlnLysLeu	1231
Q	3802	CTTAATGAAAAACGATGGAAGAAACAAGTATGATTCATCCAAAAAAGATATAGTAT	3861
D	1232	IleSerGlnGlnIleGluLysAsnAsnIleHisAsnCys-----LysLysThrGlnAsp	1249
Q	3862	GGATATCCGATTTGCCAATCCGGAATATTAATTTAGTGAAGAACCCGCGTGTATAG	3921
D	1250	Ala-----LysTrpLysCysGlnAsnThrLysLeuGlyLysAspGlnGlyValCysMet	1267
Q	3922	CCCCCTAGAAAGCAAAAGTTATGCGTCACTTTCTTGSCAAATGATTAATGAATAAAAA	3981
D	1268	ProProTyrGlyArgGlnAsnLeuCysValHisTyrLeuThrLysLeuAsnAspAspSerLys	1287

QY 3982 TTGCAATCAAGTATTTAAAGAGCTTTCATCAAAATGCGACAGCAAAACATTC 4041
 DB 1288 -----GluGluAspLeuArgGluAlaPheIleLysSerAlaIalaGluThrPhe 1304
 QY 4042 TTCTCATGTATTTATTAAGTAGAGGTGAGGAAGAAATCACTCATTAAGAAATTA 4101
 DB 1305 LeuLeuArgGluThrLysSerLysAsnValGluAsp-----AspLysIleLeu 1321
 QY 4102 AAGAAGCAAAATTCCTCCGCAATTTTGACATGCTTTCACATTTGAGATTTAT 4161
 DB 1322 HisArgAspMetLeuProGluPhePheArgSerMetPheThrPheGlyAspTyr 1341
 QY 4162 AGAGATTTTATTGGAACAGATATATCAAA-----GGTCATGTGAGGAGCAATAA 4215
 DB 1342 ArgAspIleLysLeuAspThrAspIleSerGluLysIleAlaAspHisAspValThrThr 1361
 QY 4216 CTTAAAGCAAAATAGATCTCTTTCAAAAATGTGACCAAAAATCTCTTAATGAAAA 4275
 DB 1362 AlaLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1381
 QY 4276 -----ACACGCCAAGATGTGACAGACAGATAGCATGATGATGAGCAAGCTATG 4326
 DB 1382 LysValLeuGluArgGluGluGluGluGluGluGluGluGluGluGluGluGlu 1401
 QY 4327 CTATGTGACCTA---GTAAAAATTTGGGCAAAAAGATGAT-----TTT 4368
 DB 1402 LeuGluAlaLeuSerLysThrGluThrGluLysLysMetAspIleGluValArgThrTyr 1421
 QY 4369 ACCGAAATCTACGGTTACAAACAAACGTCMAATTTAGTACAAAAGCACTTTGGAGAA 4428
 DB 1422 LeuMetLysTyrLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1438
 QY 4429 TTGCAAAAGCAAGCCGATTTTACGATGGCTAAGCAAGTATGACGACGATTTCTAT 4488
 DB 1439 PheAlaSerArgProProPheLeuArgTyrValThrGluThrGluLysAspPheValLys 1458
 QY 4489 ACACGCAAAAATATTGTAAGATGTGACAGAAAATGTAACTCA----- 4533
 DB 1459 AsnArgLysLysGluLeuValSerLeuLysLysLysLysLysLysLysLysLys 1478
 QY 4534 -----AATGACCAATTTGAAGTATGACAGAA-----TGTAATGAGAA 4572
 DB 1479 AsnGlyThrSerAsnLysThrCysAspAspAsnGluAsnCysGluAlaCysLysThrGln 1498
 QY 4573 TGGCAGACGATCGTTAAATATATGAAAAA-----AAAAACAGTGGATTCACAAAGATTA 4629
 DB 1499 CysGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1518
 QY 4630 TATTACAGAGATGACCGCAAAAAGATTTGATGACACACATTTGGTATGATT 4689
 DB 1519 LysPheGluLeuLysLysAsnSerAlaThrTyrAsnAsn-----GlyLeuAlaVal 1535
 QY 4690 ACAGATTAATGACAGCAAGTACAGATTAATTAACAGAAATTTACTGTAATGTG 4749
 DB 1536 LysGlu-----AlaAsnSerGluThrTyrLysAsnAspProGluValThrGluAla 1552
 QY 4750 GGTGATAGCCTGGAAGTGCCTGTGTGACAA-----AGAATATACAAATTTGTTA 4800
 DB 1553 AsnSerAlaLysHisAlaArgAspTyrLeuLysThrGlnLeuLysMetIleCysThr 1572
 QY 4801 GAAAAACAGGCTTACTATGATGCCGACAAACATTTGGGTGACAAAATTTATGAAAT 4860
 DB 1573 AsnGlyAsnThrTyrLysAsnCysAspTyrThrCysMetAsnThrSerSerThrAsn 1592
 QY 4861 GAGCAGC-----AAATATACAAATTTGAGTAAATATGAGTAAAGCA 4905
 DB 1593 SerGluMetProAlaSerLeuAspTyrThrProSerGluTyrLysAspLysCysAsn 1612
 QY 4906 TTAGTAAGAGGCAAGC 4923
 DB 1613 ValProAspIleCysSer 1618

RESULT 11
 ID 08T5G0 PRELIMINARY; PRT: 2658 AA.
 AC 08T5G0;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Erythrocyte membrane protein-1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=JDP8;
 RA Chactopadhyay R., Pillai C.R., Chitnis C.;
 RT "Identification of a domain responsible for binding to intercellular
 RT adhesion molecule-1 from a Plasmodium falciparum field isolate."
 RL Submitted (Mar-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY028643; NAK49742.1; -
 FT NON_TER 2658 2658
 SQ SEQUENCE 2658 AA; 301326 MW; AEE33BC9167E7B7 CRC64;
 Alignment Scores:
 Pred. No.: 8.7e-137 Length: 2658
 Score: 2869.00 Matches: 872
 Percent Similarity: 38.52% Conservative: 373
 Best Local Similarity: 26.98% Mismatches: 909
 Query Match: 14.75% Indels: 1078
 DB: 5 Gaps: 122
 US-10-087-013-1 (1-10628) x 08T5G0 (1-2658)
 QY 91 AGTCAAAAGTCAAGAAATGTTTGAAGCTTATGCCAAATATTAAGATCATCA 150
 DB 11 SerGluGluThrAlaLysGluLeuLeuAspArgIleGlyGluGluVal----- 26
 QY 151 AATATGCAAAAGACATGTGATTCGTTGAAGGGAT-----TTG 192
 DB 27 ---TyrGluLysAlaHisGluAspAlaLeuLysGlySerMetGlyLeuLysGlyThrLeu 45
 QY 193 ACGAAAGCAAAATTCGTCGTGTCCTTCACCCAGTAATAAGCATATTTATAT 252
 DB 46 SerGluAlaIlePheGluLysAlaProGluGlyLysGlnThrSerGluAsp----- 62
 QY 253 CCATATCATGTAATTTAGATCATTAAGAACATTAATTTACG-----TATGATGAT 306
 DB 63 -----ProCysAspLeuAsnHisGluTyrHisThrThrValThrSerGlyTyrAspLys 80
 QY 307 GTGAATTTGAGACATCCTTGCCATGGTAGAACAACCGATTTGATGAAGTGAAGA 366
 DB 81 GluAsn-----ProCysLysAspArgProGluValArgPheSerTyrThrGluGly 97
 QY 367 TCTGATGTGGA---AATAAATACGTAATTTAAAAAGAAAAATATGCTATAGCTGT 423
 DB 98 AlaGluCysAspLysSerLysIleArgGlySerAsnSerAsnLysAspGly---AlaCys 116
 QY 424 GCGCCACATGAGACAGCATATGTGTGATAAAGCTTGAAGCTGTA---AATGATATA 480
 DB 117 AlaProPheArgArgLeuHisLysLysCysAspGlnHisLeuGlnIleLysHisAspLys 136
 QY 481 AATACCAAAATATGATATTTATGGGAATGTAAGTACTAGTACACCAAAATGAGAGT 540
 DB 137 IleThrArg-----HisAsnLeuLeuAlaAspValCysGluAlaAlaLysPheGluAla 154
 QY 541 GATTCATTT-----GTAAATATGATCCACATAAAGAACT 576
 DB 155 GluSerLeuGluLysTyrArgGlyGlnTyrGlnLeuAsnAsn----- 168
 QY 577 TCAGACGCT-----TGTAAGCTCTTGCAGAGAGTTCGACAGATATAGCTGAT 624
 DB 169 SerAspValAsnIleAsnIleCysThrGlnLeuAlaArgSerPheAlaAspIleGlyAsp 188

QY 625 ATTGTAAGAGAAATGATATGTTTAAACAAATGCTCATGACAAA----- 669
Db 189 ILevalrvglyLysAspLeuPheLeuGlyAsnThrTyrgLysSerAlaGlnArgLle 208
QY 670 GTAGAAACGGGTCTCCGAGAGGTTTTCAGAAATATACATGAT-----GGA 714
Db 209 LeuGlnAsnLysLeuLysGlnLlePheGlnLysLleHisSerAspValThrLysGlyGly 228
QY 715 ATGCAAGATCAAGTAAATAAATGATTACAAATCGATGATCTGCAAAATTTATTAATTA 774
Db 229 ThrAsnGlyLleuLysLysArgTyrgLulLysAspLylGln--AsnPheTyrgLln 247
QY 775 AGAAGACATGTTGGAATGTGATGAAATAAATGGAATGGAACCTATACATGTGATGCA 834
Db 248 ArgGlnAspTrpThrPheAlaAsnArgLulThrValTrpTyrgLalThrCysGlyAla 267
QY 835 TCATATTAATCTGGAATTTTATGCAATCAGAAGTAATACCACTATATTTGCAAAATCC 894
Db 268 GlyThrSerAspLysTyrgPheArgAsnThrCysSerAsnAspLysAlaGlyThrSerGly 287
QY 895 AAATGGCGCATTAACAGAGAAAGTTCTTACCAATTTAGATTTATGCTCCCAATATTTA 954
Db 288 LysCysArgGlyAsnAspAsnGlnValProThrTyrgPheAspTyrgValProGlnTyrgL 307
QY 955 CGTTGGTTCAGCAATGGGAGAGAGTTTGGCGAAAGAAATTTAAATTTGAAAG 1014
Db 308 ArgTrpPheGlnLulThrPheAlaGlnAspPheCysArgLysLysHisLysLeuGlnAsn 327
QY 1015 GTCAAGACTCTGTGCT-----AATGACAAAGAACGCTTATATGTAGTATATGCA 1068
Db 328 AlaLysAsnLysCysArgLysProAsnGlyValGlnLys--TyrCysAspLeuAsnArg 346
QY 1069 CATGATTTGACGACACTTTTGGAAATAAGATTTTTCATTTGGATTAATAGTGTACT 1128
Db 347 TyrAspCysGlnLulThrAlaSerGlyLysHisAspPheGlnAspAspValCysLys 366
QY 1129 GACTGTTCACATAAATGCAAGTTTGGATTTGGTGAATGGAATCAACATGATTT 1188
Db 367 AspCysGlnLysSerCysAlaProPheValAspTrpLleAspAsnGlnLysLeuLphe 386
QY 1189 AAAAAACAAGAAAGAAATTAAGAAATGAAATACATCATATTTTCGACGATTAACA 1248
Db 387 LeuLysGlnArgAsnLysTyrgLulLysGlnMetGlnLysTyrg--ThrAsnGlyThrLys 405
QY 1249 TTTGTCAATATATTAATAGTGA----- 1272
Db 406 -----AsnSerLysArgLysLysProGlnAlaGlyLysSerAsnTyrgAsp 420
QY 1273 ---TATTAACAATTTTATGAAAAACTTAAGAAACGCAATATGCAATATGACACT 1329
Db 421 GlyTyrgLulLysArgPheTyrgLlnLleLeu--GlnGlyLysTyrgAsnValAspLys 439
QY 1330 TTTTAAATTTACTTAATGAAGAAAGATATGTAA-----GGA 1368
Db 440 PheLeuAspLeuLysAsnAspLulThrThrCysLysLysAsnAsnGlnLleGlnLulGly 459
QY 1369 GGATTAACAGAGAAAGATATTAATTTTACTTAACAGTGTGATGAC-----AAA 1419
Db 460 GlyLlnLleAsnPheLysAsnValAsnSerGlyLysAsnSerAspAspAspSerAsn 479
QY 1420 GGGATATTTTTCGTTTCGAATATGCGCAAGTGTCCGCACTCGGGGTCAAAATGTAT 1479
Db 480 LysThrPheCysArgThrThrTyrgCysGlnAlaCysProTrpCysGlyAlaGlnLulAsp 499
QY 1480 GGTATTAATATC---ACACACAAATCAGATATGATGATGCTGACGTAATAATGAAGAC 1536
Db 500 AsnSerLysSerGlyLysTrpLysAlaLysAsnAspAlaAspCysGlyLysLysLysAsp 519
QY 1537 TATTAACCTTCATGGGTGAGACCTACTATATATCACTGTCTTATAGTGAAATGAA 1596
Db 520 TyrAspProGln-----LysThrThrLleGlnLleuThrGlnLysAspThrArg 536
QY 1597 CAAGGTGATATTAACAATAATTTTGTGACAGCTCACT----- 1644

Db 537 LysSerAspMetValGlnLysTyrgLysLysPheCysThrSerAlaThrGlyAlaThrGly 556
QY 1645 -----AATTAACAA----- 1653
Db 557 AlaProAspThrAlaThrGlyGlnAsnGlyLysLysGlyAlaSerGlyLysAsnGly 576
QY 1654 GATTAATAATCAAAAAATGGAATGCTATTATTAAGATGAAT----- 1698
Db 577 AspAsnLleThrGlnThrPheThrPheCysTyrgTyrgLysLysAsnGlyLysAspValAl 596
QY 1699 -----ATTAATGATGTAACTGGAACAAATAATGCAATCAATCAATGATATCT 1749
Db 597 LysLysAspLleAsnPheCysValLeuGlnAspGlyLysGln--HisThrLysGln 615
QY 1750 AAGATATATCACTTTTCATTAATTTTTCGAATTAAGGTTACATATTTAAGGATCT 1809
Db 616 LysValThrSerTyrgAsnValPhePheTrpLysTrpValTyrAspMetLleHisSer 635
QY 1810 ATTAAGTGAATGACAAACTTAAACTGTATTAATAATACA---ACCAGCATTTATTT 1866
Db 636 IleLysTrpArgAsnGlnLeuArgSerCysLleAsnAsnAlaLysSerGlnAsnCysLys 655
QY 1867 GAT-----GAATGTAACAGAAATGCTTATGTTTTCAGAGATGAGTTAAACAAAGAA 1920
Db 656 AsnAsnLysLysCysAsnSerAspCysGlyCysPheLysGlnTrpLleGlnLulLysArg 675
QY 1921 GAAGATGATGATGATTAACAAACTGTC--ACAAAAAAGAAATATACAGCAATCG 1977
Db 676 LysGlnTrpAspLysLleLysAspHisPheLysThrGlnAspPheGlyLleProGlyGly 695
QY 1978 TATTAATTAATTAATTAATTAATTTTGAAGTATTTTGAATTTTGAATGATTAAGT 2037
Db 696 ProLeuGlnLlnPheAsp-----TyrAspPheValLeuLysPheValLeu 710
QY 2038 GACAAAGATGAGCAAAATGGAAGAACTTATGGAATAATTAATAAAGAAAAATGAG 2097
Db 711 AspLys-----LysGlnLeuLeuLlnAsnLleLysAspThrHisAlaAsp 725
QY 2098 TTTTCCATTTGAAATAATATAGGACTATTTTACAGAAATGCA----- 2139
Db 726 AlaLysAspLleGlyAlaGlyLleGlnLysMetLeuGlnAlaGlyAlaAspAlaSerGly 745
QY 2140 -----ATGAACTCTTGTTA 2154
Db 746 SerGlyGlyGlyAlaGlnGlyAlaLysGlyLysHisAsnThrLysLleAspLysPheLeu 765
QY 2155 GATCACTTAAGAAAGAACTGCCAGCATTTGTAAGACATATATCAAAACGAAGCATGTGAA 2214
Db 766 GlnGlnLulGlnGlnPheAlaGlnLulThrCysLys-----GlnThrGlnAspAsnCysPro 783
QY 2215 ACATCCCATATGACAAACAAACCCGTTGTTAAACCTCGTGAGGACGCAACCACT 2274
Db 784 LysLysProThrLysValAlaArgAsnProCysTyrg-----GlnAsnAsnThrTyrgAsp 800
QY 2275 AAAAATTAAGAAATAGCAACAATCTTAAAGAGATGCAATACAGAGACGCAAAAT 2334
Db 801 AlaLeuAlaGlnLulLysValAlaGlnLleLeuGlnGlyAlaGlnLlnThrGlnLleuLleAsn 820
QY 2335 ---CGTGTCTCATTAATGAAGAAAGAAAGGCAAGGATATATTAACGTGGGGGT 2391
Db 821 HisArgSerLysSerLeuLysGlyAspProGlnGlnGlyHisTyrgAsnGlyAsnGly 840
QY 2392 AGGAAAGACCTTCAAGACAAATTTATGATATATATATTAACATCTTAATGATAT 2451
Db 841 AsnLysSerValLeuLysAsp--ValCysLysLleThrAsnGlnLysSerAsnAlaGly 859
QY 2452 CTGGGTTTTCAATGACCATGTGATGCAAGCAAGCAAGGTGATGATTAACAACAA 2511
Db 860 LysAsnLysSerAsnAspProCysAsnGlyLysAsn-----GlnAsnArg 874
QY 2512 TTTGTGTAAGTGAATGAAGGAAATGATCGGAACACATGCGTAAAGATACAGAAAT 2571
Db 2512 ----- 2571

Db 875 PheAsnIleGlyGluLysTrpLysAsnGlyGluValLysMetSerAspThrHisSer 894
QY 2572 GTTATTATGCGTCCAGAGAGACATATATGATCCCAATTTGGACATTATACAAAGC 2631
Db 895 ---TyrMetProProArgTrgIuHisPheCysThrSerAsnLeuGluHisLys 913
QY 2632 GATGATCACCCCTTAATGTAATGTTGATTTGATTTGTTAATTCCTTTTGGGG 2691
Db 914 AspLysGlyGlyArgPheGluGluValProAspAsnLysAlaThrHisSerLeuLeuGly 933
QY 2692 GATGTTCTCTTATGCAAAATATGAGCAAAACAGATATATGCAATATGATTAAGAAAG 2751
Db 934 AspValLeuIleAlaLysLysGluAlaGluLysIleLysGluLeuLysTrgIuLysAsn 953
QY 2752 AATTAACCTTAAGGCCCAAGAGTAAGTCAACCAACACAGACAACTATCTGTCA 2811
Db 954 LysAspGlnSerGlyLysAsnAlaLysAsnGlyLeuAsnAsnGluLysThrValCysArg 973
QY 2812 GGTATTAAGTTCATGTTTCCAGATATAGTATATATGATTTGAGGAAAGATCTGTGGAA 2871
Db 974 AlaIleArgTrpSerPheAlaAspIleGlyAspIleIleLysGlyThrAspLeuTrpAsp 993
QY 2872 AGAAGCGGAGCATGGTAAAGCTGCAGAGACATTTGGAACCTGTTTGTGATATACAT 2931
Db 994 GlnAsnLysGlyGluThrAspThrGlnSerAsnLeuValThrIlePheGlyLysIle 1012
QY 2932 AAGTCATCAAGCAAGCAAGAAATGAT---AATATATATGATGATGCCCCCAATAT 2985
Db 1013 ---LysGlyThrLeuAsnAspTrpSerLysTyrAsnAspLys---LysHis 1027
QY 2986 TTAATAATGAGGAAATGTTGGGAGGATATAGCCCAAAATGAGGAAAGCAATGAGAA 3045
Db 1028 LeuGluLeuArgLysAspTrpTrpGluAlaAsnArgAlaLysValTrpGluAlaMetLys 1047
QY 3046 TGTGATATATAATATTATGAGATAAATCGGACACCAATCAACAAGTATGTTATTCG 3105
Db 1048 CysAlaThrLysAspIleProAspMetLys----- 1057
QY 3106 GGATATAGTATCATACACCATGATGATATATATCCCAAAATTAAGATGATGACG 3165
Db 1058 ---CysAsnGlyIleProIleGluAspTrpLysIleProGlnArgLeuArgTrpMetThr 1075
QY 3166 GAATGGCAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3225
Db 1076 GluTrpAlaGluTrpPheCysLysAlaGlnSerGlnLysTrpAspLysLeuGlnTrgIu 1095
QY 3226 TGTAAAGAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3285
Db 1096 CysGlyIleCysLysLysLys-----ValGlnGlyCysThr-----SerGlyAspGly 1111
QY 3286 ---TGTACGAGTGCACAGACCTTGTAAATGATATATGATATATGATGATGATGATG 3342
Db 1112 LysCysGluLysCysGluAlaLysCysGluAlaLysTrpLysThrLysIleAspThrTrpLys 1131
QY 3343 GAACATGATGAT 3402
Db 1132 GlnGlnTrpAspLysMetCylLeuLysTrpLysLeuLeuTrpLysAlaAsnAlaLysThrTrp 1151
QY 3403 GTTATGATATAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3462
Db 1152 SerThrAsnAlaGlyArgThrValLeuGlyAspAlaSerProAspTrp---GlnGlnMet 1170
QY 3463 ATTTGAATTTTGTCCGAAATTAACAAACA-----AATGGTGCACAA 3504
Db 1171 LeuAspPhePheLysLysLeuGlnLysGluIleLysAsnSerAlaLeuAsnArgTrpPhe 1190
QY 3505 AGTAATAAAGAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3564
Db 1191 ArgSerIleAspGlyLysAsnAsnAspProIlePhe-----ThrSerProLysSer 1207
QY 3565 AATGTTGAGCATATCTCATGATATACAGAAATTTGATGATGATGATGATGATGATGATG 3624
Db 1208 SerAlaGluGlyLysTrpIleHisGlnGluLeuProHisThrGlnCysAspAlaGlnLysTrp 1227

QY 3625 TTTTGTGATGAAAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 3684
Db 1228 PheCysAsnThrAsnLys-----AsnLysGluLysTyrAlaPheArgHisGlnPro 1244
QY 3685 CAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3732
Db 1245 HisAspTrpAspGluAlaLysAlaCysArgProProSerThrProGluSerProAlaArg 1264
QY 3733 ---AGGTACAGATTAATAACGAAA----- 3753
Db 1265 AsnLeuProAlaAlaAspGlyGlyValGluGluAlaGluLysGluGluGlu 1284
QY 3754 ---AAAAAGCGGAGAAAGAT 3774
Db 1285 GluGluAspValGluLysAspAspAspAspAspAspAspAspAspAspAspAspAsp 1304
QY 3775 ACGGATGTAATAACAGTGAATGATATATATATATATATATATATATATATATATATAT 3834
Db 1305 ProGlnCysLysThrValAsnAspIleLeuSerThrAspAspArgTrpLysGlnValGly 1324
QY 3835 GATTTGCATCCAAA---AAGATAGTAAT---GATATCCGATGCGCAATGCGGAAAT 3888
Db 1325 GlnCysAsnAlaLysIleLysAsnIleAsnGlnSerTrpProAspTrpThrCysValAsn 1344
QY 3889 ATTAATTAAGTGAAGACCT---CGTGTGTATGCCCCCTTGAAGACAAAGTTATGCT 3945
Db 1345 SerLysPheGluAsnAsnGlnLysTrpGlyProCysMetProProArgTrgIuLysLeuCys 1364
QY 3946 GTACATTTCTTG-----GCAATGATTAATAATAATAATAATAATAATAATAATAACA 3993
Db 1365 LeuTrpTrpLeuLysGluLeuGlyGluAsnAspAspGlnLys----- 1379
QY 3994 GTTAAATTAAGACACTTTCATCAATCTGCAGCAGCAAAACATTTCTTCATGATGAT 4053
Db 1380 ---PheLysAspAlaIleIleLysIleLysIleLysIleLysIleLysIleLysIleLys 1397
QY 4054 TATTATAAAGATGAAGTGTGAAGAAATGAATGAT---AAAGATTAATAAAGAGGC 4110
Db 1398 TyrTrpLysSerLys-----AsnSerMetAspIleLysLysLeuGlnSerGly 1413
QY 4111 AAAATCTCCCGCATTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 4170
Db 1414 GluIleProGluLysLeuPheLeuArgSerMetLysPheThrTrpLysTrpAlaGlyAspIle 1433
QY 4171 TTATTGGAACAGATATATCAAAAGCTCATGTCGAGAGAAATTAATAAAGAGCAATA 4230
Db 1434 CysLeuAsnThrAspIleSerLysLysGlnLysAlaValSerAspAlaLysGlyLysIle 1453
QY 4231 GATTCCTTTTCAAAAAGGTGACCAAAAATCTCTAATGAAAAAACGCCAAGATGG 4290
Db 1454 AspAlaLysPheAsnLysTrpThrAspThrAsn-----ArgThrLysTrp 1468
QY 4291 TGGACAGAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4341
Db 1469 TrpAspThrAsnLysProGluIleTrpGluGlyMetLeuCysAlaLeuThrHisGlyVal 1488
QY 4342 AAAATGGGGCAAAAAGATGATTTTACCAAAACATGAGTGAACAACGTCAAATTT 4401
Db 1489 ThrAsnThrAspAsnLysArgLysIleLysThrAspLysSerTrpLysGluLeuGlnSer 1508
QY 4402 AGTGACAAAAGCACACTTTGAGAGATTTGCCAAACGACCCAGTTTTCAGATGGCTA 4461
Db 1509 LysAsnValThrThrProLeuGlnLysPheAlaGlnArgProGlnPheLeuArgTrpPhe 1528
QY 4462 ACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 4521
Db 1529 IleGluTrpGlyAspAspPheCysArgGluGlnLysLysTrpArgGlnGluLysGln 1548
QY 4522 AATGTAAG-----TCAATGACCAATTAAGTGTGAT--- 4554
Db 1549 LysCysAsnLysCysCysAsnAsnGlnLysAsnValThrSerAspLysLysThrLysCys 1568

Db 1921 GlypheLeuArgGlnMetPheTyrThrLeuAlaAspTyrLysAspIleLeuTyrSerGly 1940
Qy 6634 GATATTTCATGATATAAAATATA----- 6660
Db 1941 SerAsnAspAsnLeuLysHisIleValLeuGluAlaSerGlyThrLysGluLysGly 1960
Qy 6661 -----ACGTACCAATAGCTGACACACCATCTCTCAACGAAATATAAAGAAAAA 6711
Db 1961 AspMetGlnLysIleGlnAspLysIleLysLysThrLeuAsnGlyAspAsnAsnGlnGlu 1980
Qy 6712 CAG-----GATAAAAAAAGATGAAGAAATACGTAATATTTGG 6753
Db 1981 SerGlyHisProProSerProSerGlyLysAsnSerValThrThrProGlnThrPTr 2000
Qy 6754 GAGAAATATAAAATTTATTTGGAGAGATGATATGATTAATCTATCTCACA 6813
Db 2001 AsnAlaHisAlaProSerIleThrHisGlyMetIleArgAlaLeuThrHisLysGlnSer 2020
Qy 6814 GAC----- 6816
Db 2021 AspAlaThrIleAlaGlyLysLysIleGlnGlnAsnThrGlnLeuLysAsnAlaLeu 2040
Qy 6817 -----GAAAGCAAAAGAAAGAAATTAAGATTAATACAGATGACATGACCAACTG 6873
Db 2041 LeuAspAspThrLysAsnLysProLysAspAsnTyrLysTyr-----AspLysValLysLeu 2059
Qy 6874 -----ACGCTTCCTT 6885
Db 2060 AspGluAsnSerGlyThrSerProLysProAlaGlyValAsnGlnAlaProProLysLeu 2079
Qy 6886 GAAAGATTTGTAAGAAAGCCCAATTTTGACATGCTGACAGATGCGCAAGAAATTT 6945
Db 2080 ThrGlnPheValGlnLysProProLysPheValGlnLysGlnLysLysLysLysPhe 2099
Qy 6946 TGTAAATAGAGAGAAAGTGTGTAATTAATGAGCGGCTGTAG----- 6993
Db 2100 CysArgLysGlnLysHisLysLeuGluIleIleArgValAspCysArgGlyLysAspGlu 2119
Qy 6994 GAATATGAGTGAATGTAAT----- 7017
Db 2120 AspLysHisCysSerGlyTyrGlyLysAsnCysAspAspAsnLeuIleAlaAspProSer 2139
Qy 7018 -----GACGTAAGACACAGAAATGCGAGCGCTGTACATATCAAAATTTT 7068
Db 2140 IlePheProAspLeuAsnCysProGlyCysAlaLysHisCysSerGlyTyrLysTyr 2159
Qy 7069 ATTAAGAAAGTGAAGAAATGATATGAAAGCAAGAAAGTCAAAAGATTAAGAT 7128
Db 2160 IleGlnArgLysLysIleGlnPheThrGlnGlnAspAsnAlaLysAsn----- 2176
Qy 7129 GGCAGAAAGTATAGATTTATCTCTACTGAAAGACATGAGAGCGCAATGTGCT 7188
Db 2177 -----GlnLysValAsnCysGlu 2182
Qy 7189 CATGATATTTTAACATGAATTAAGAATTTGCGCAATAAGATTTCTTGTATG 7248
Db 2183 LysGln-----SerLysGlyLysAsnGlyValCysGlyLysLeu 2196
Qy 7249 CAAGAACTCTCTCACACTACCAAAACACAAACATCAACATTCGATGTAT 7308
Db 2197 Gln----- 2197
Qy 7309 GATATGCCAGATTCCTGCTATTTGTTCTGAGAAATTTAAACAGTGTGATGCTGAA 7368
Db 2197 ----- 2197
Qy 7369 CTTTCAAAAAGGATCTATGATTCATACAAAAAATTACGACCTAAATACGTATG 7428
Db 2197 ----- 2197
Qy 7429 AATGTGTAGAGAAACAGCATATTTATCTTAAGAA-----GCAGAAATATAT 7479
Db 2198 -----GlnAsnAlaAlaLysPheLeuGlnLysLeuGlnLysLeuGlnLysSerLysAsnAsn 2214

Qy 7480 ARGAT-----ATTACCTTGACGAAAAATTTATVACCTATGTGCTTACA 7524
Db 2215 LysAspAsnGlyAspLysThrIleAsnPheLysGlnProAspValThrPheLys----- 2232
Qy 7525 AAGCAAAAGAAAGTAAAAATAGTTGACATTAATATACCTTCGATCTCAAGAACCT 7584
Db 2233 -----ProLysAsnCysLys--- 2238
Qy 7585 TATGACCTGATTAATATATATAGAGAGAAACCTTGCGAAATATAGAGAAATCGT 7644
Db 2239 -----ProCys-----SerGlu 2242
Qy 7645 TTTAAGTATGATATGAAATGAAATGTTCACAAAATTCAAAGTTCTATACGAGAAATA 7704
Db 2243 PheLysIleAsp-----CysThrLysAlaLysCysThrGlyAspGlnGluLys 2258
Qy 7705 AGAGTATGTGTACCTCCAGAGAGAAACATATGTGCTTAAGAAATTTAGTGAATTTAA 7764
Db 2259 ArgLysCys-----AsnValLysAsnGlnThrValIleArg 2270
Qy 7765 ATGAAAGACTTAAGATAGTATTAATTCCTTAAAAATGTTGCTGAACTGCAGCAAT 7824
Db 2271 AlaThrAspLysAspAspLys-----AsnGlyAsn 2281
Qy 7825 GAGCAATAGACATTAATTAATAAACTTCACCTCAGAGACGGTGCCTAATGAATCCAATA 7884
Db 2282 GluAsnIleAsnMetVal-----Val 2288
Qy 7885 TGTGATACATGAAATATATAGTTTCGCTGATCTGGTGACATACTTACAGACAGATATG 7944
Db 2289 SerAspThrSerLys-----LysGlyAsp-----GlnAspAsp 2299
Qy 7945 TTACGAATGCTGGTACTTACCTCCCGTAGAAATTAATATAGGTTTGAATAC 8004
Db 2300 LeuArgVal----- 2302
Qy 8005 ATATATGAGAAATGAGAAATTAATAAGTAGAAATTAATACACATGTACAAACG 8064
Db 2302 ----- 2302
Qy 8065 TTTGCTTCCTGCTGGTGGATGCTATAGAAAGATTTTGGAAACCATGACGTGCAAA 8124
Db 2303 -----CysArg 2304
Qy 8125 GCACAGAGATGCAAAACTTTTGAAGAAAGAGATGATGATGATGAAACCATACACA 8184
Db 2305 -----AspAlaGlyIlePheLysGlyIleArgLysAspLysPhe----- 2317
Qy 8185 TTTATACAGATTAAGTGTGACATTAAGACATGCACCTGTT----- 8226
Db 2318 IleCysGlyAsnValCysGlyTyrAsnValCysLysAspValLysValAsnGlnSer 2337
Qy 8227 -----GATGATTAATATACCTCAACGTTTCGATGATGATGCTGAATGATTAAT 8280
Db 2338 GlyAspGlyAsnGlnIleIleIleLysAlaPheLysIleThrLeuAlaTyrPhe 2357
Qy 8281 TGTAAAGCACTGATGAGAAATGGAATAATTTAAATAATCATGATGATCACTG---AAA 8337
Db 2358 -----LeuGlnAspTyrAsnLysIleLysLysLysLeuLysSerCysThrLys 2373
Qy 8338 ACATCTGACAGATGCAAGATGATTTATGATGAAATAATAGTGTACAGCTTAACAGACA 8397
Db 2374 SerSerAsp-----AlaThrProCysIleLysGly 2383
Qy 8398 TGT-----CAAGATATATAAAATTTTGTCTTAATGGAATCTGATTCGATATACAA 8451
Db 2384 CysAlaLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2403
Qy 8452 TCAATTAATAC-----AAAGAAATGATATGACACACCAATATATACA 8493
Db 2404 LeuGlnGlnLysGlnLysLysSerSerAspLysSerPheLeuValLysThrIleLeuGln 2423

QY	937	TATGCCCTCAATATTTAGCGTTGGTTCGACGAAATGGGGAGAAAGACTTTTGCCGAAAAAGA	996
Db	81	TyTAlProclnTyLeuAlgtPrPhesPrGlnPrGlnGlnPrHeCySaGlyLysArg	100
QY	997	AATTTAAATGAAAAAGGTCAAGAGCTCGTCTTAATGACAAAGAACTTATATGT	105
Db	101	AsnIleLysLeuLysLysValLysAspSerGlyArgAsnAspLysGlnAlaGlyLeuTyCys	120
QY	1057	AGCATAAATGACATGATTTGTGCGCAACCTATTGGAAAAAGGTATTTGCAATTGGAT	111
Db	121	SenHisAspLysTyLysAspCysThrThrThrIleTyLysGlySerLeuHisLeuAsp	140
QY	1117	AATTAAGTACTGACGTTCGACATAATGCAAAAGCTTTTGAAGCTTTGGTAAGGAATGCA	117
Db	141	AsnLysCysGlyThrAspCysLeuThrLysCysLysValAlaPheGlnValAlaPheGlnGln	160
QY	1177	CAAGAAGCATTTTAAAAACAAAAAGAAAAATATGAAAAAGAAATACATCATATTATGCG	123
Db	161	GlnGlnAlaPheLysLysGlnLysGlnLysTyGlnLysGlnLysGlnLysGlnLysGlnLysSer	180
QY	1237	AACGATACAAATTTGTCAATATATTTTAATGTGCAATATTAACATTTTATGAAAA	129
Db	181	AsnAspAlaLysPheValAlaAsnAsnIleAsnSerGlnTyTyLysGlnPheTyGlnLys	200
QY	1297	CTTAAGGAAGCCAAATATGCAACTATATGACACTTTTAAATTTACTAAATGAAAGAAAG	135
Db	201	LeuLysGlnThrGlnTyTAlAlaThrAsnAspPheLysLeuAsnLeuAsnGlnGlnLys	220
QY	1357	TATTGAAAGAGGATTTACGACGAGGAAAAAGATATTTCTTTTACAGCTGCTATGAC	141
Db	221	TyTyCysLysGlnGlyLeuProGlnGlnLysAspIlePhePheThrAsnSerAlaAspAsp	240
QY	1417	AAAGGATATTTTATCGTTCCGAATATTTGCCAAGTGTCGCCGACGCGGGCTCAAAATG	147
Db	241	LysGlnIlePheTyArgSerGlnTyTyCysGlnValCysProAspCysLysValLysCys	260
QY	1477	GATGGTATTAATTCACACACAAATTCAGATATATGATTCGGAACGTGTAAATATGAAGAC	1536
Db	261	AspGlnIleLysTyTyThrHisLysSerAspAsnAspArgGlnAlaValAlaAsnGlnLysAsp	280
QY	1537	TATTAACCTCCATGGGGTGTGAAGCCTACAAATATACAGTCCCTTATATGATGTTAAATA	1596
Db	281	TyTyLysProProTrpGlnLysValLysProThrAsnIlePheValLeuLysSerGlnLysAsnGln	300
QY	1597	CAAGGTATATTAACAAAAATTTGAAAAATTTTGTAAACGTCACATATTAACAAGAT	1656
Db	301	GlnGlnLysPheIleThrGlnLysLysGlnAsnPheCysAsnSerSerThrAsnTyLysAsp	320
QY	1657	AAAAATATCAAAAAATGGGAATCGTATTAAAGGATGAAGAAATATAATATGATGTAACTG	1716
Db	321	LysAsnAsnGlnLysTrpGlnCysTyTyTyLysAspGlnAsnIleAsnArgCysLysLeu	340
QY	1717	GACAAAAATCTGAAATCAATTAATGATTAATCCTAAGATATATCATTTCAATATTTT	1776
Db	341	GlnGlnAsnThrGlnIleAsnAsnAspAsnProLysIleIleSerPheHisAsnPhePhe	360
QY	1777	GAATTAAGGGTACATATTTATTAAGGATCTATTAAAGTGGATGACAACTTAAACT	1836
Db	361	GlnLeuTrpAlaIleThrTyLeuLeuAlaGlyAspThrIleLysTrpAsnAspLysLeuLysThr	380
QY	1837	TGTATAATATACACACACACGACTGTATTGATGCAATGATTAACAGAAATGGCTATGTTT	1896
Db	381	CysIleAsnAsnThrThrThrHisCysIleAspGlnCysAsnAlaGlyAsnCysLeuCysPhe	400
QY	1897	GACGATAGGGTATAACAAAAAGAAAGAAATGGAATAGTATTAAGAAACCTGTTACAAAA	1956
Db	401	AspArgTrpAlaLysGlnLysGlnLysGlnGlnGlnTyTrpAsnSerIleLysGlnPheThrLys	420
QY	1957	AAAAAGATTTAAGCAATCGTATTATAGTATATATATATATCTTTTGAAGGTATATTT	2016
Db	421	LysLysAsnArgGlnGlnSerTyTyTySerAsnIleAsnAsnLeuPheGlnGlnTyTyTyPhe	440
QY	2017	TTTTAAAGTTATGATTAACCTTGACAAACATGACGAAAAATGGAAGAACTTATGGAAGAT	2076

```

Db 441 PheLysValMetLysPheLysLeuAspLysGluAlaLysTrpLysGluLeuMetGluAsn 460
Qy 2077 ATAAAAAGAAAAAAAGAGTTTTCATTTGCAATTTGGAATAATAGGACATATTAGAGAAAT 2136
Db 461 IleLysArgLysLysAsnGluPheSerAsnLeuGluAsnArgAspTrpLeuGluAsn 480
Qy 2137 GCATATGAACTCTGTAGATCACTTAAAGAAAGCCACGATATGATTAAGCAATTAAT 2196
Db 481 AlaIleGluLeuLeuLeuAspHisLeuLysGluThrAlaThrIleCysLysAspAsn 500
Qy 2197 ACAAGCAAGCATGTGAACATCCCATTAAT 2226
Db 501 ThrAsnGluAlaCysGluThrSerHisAsn 510

RESULT 13
P90580 PRELIMINARY; PRT: 2647 AA:
P90580:
01-MAY-1997 (TREMBLrel. 03, Created)
01-MAY-1997 (TREMBLrel. 03, Last sequence update)
01-DEC-2001 (TREMBLrel. 19, Last annotation update)
FCR3-VART11-1 protein (Fragment).
DE FCR3-VART11-1.
GN Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=97154495; PubMed=9001213;
RA Hernandez-Rivas R., Mattei D., Sterkers Y., Peterson D.S.,
  Wellem's T.E., Scherf A.;
RT "Expressed var genes are found in Plasmodium falciparum subtelomeric
  regions."
RL Mol. Cell. Biol. 17:604-611(1997).
DR EMBL: U67959; AAC47438.1;
DR InterPro: IPR004258; PFEMP.
DR Pfam: PF03011; PFEMP. 2.
FT NON_TER 2647 2647
SQ SEQUENCE 2647 AA; 303263 MW; 65FD700611348BC7 CRC64;

Alignment Scores:
Pred. No.: 1,22e-129 Length: 2647
Score: 2727.00 Matches: 922
Percent Similarity: 34.82% Conservative: 399
Best Local Similarity: 24.30% Mismatches: 810
Every Match: 14.02% Indels: 1664
Gaps: 129

US-10-087-013-1 (1-10628) x P90580 (1-2647)
Qy 91 AGTCAACAAAGTCGCAAGAAATGTTTGGACGTTATGCCAAAATATTAAGACATCATCA 150
Db 9 ThrAsnLysSerLalLysGluValLeuAspGluIleGlyLuhrrIleGln 25
Qy 151 AAATATGCAAAACATGTGATTCG-----TTGAAAGGGGATTTG 192
Db 26 -----LysLysAlaHisSerAspAlaAspThrPheArgSerGlnLeuLysGlyAsn 43
Qy 193 ACGAAGCAGAAATTCGTGTGTGT-----CCTCTACGCCAGTAATTAAG 237
Db 44 GlyLulAlaLysPheArgLysGluGlyLulIleMetGlnTrpAsnSerLysLys 61
Qy 238 CATATTAATTAATTCATATTCATTAATTTAGATCATTAAGACATATTAATTAACG 297
Db 62 -----CysGluLeuAspHisThrIleAspThrAsnValThr 73
Qy 298 TATGATGATGTGATTTGAGACATCTTGGCATGTAGAGACAAAACGATTTGATGAA 357
Db 74 AspGlyHisSerAsn-----ProCysGluGlyArgGlnThrValArgPheProAsp 90
Qy 358 GATGAAGAATTCGATGT-----GGAATTAATAATGTAATTAATAAGAAAAAATGATGCT 414

```

```

Db 91 AspAsnArgSerGlnCysThrLysAsnArgGluLeuAsp-----SerValAspAsnSerVal 109
Qy 415 ATAGCCCTGTCGCCACACCTAGAACACATATGTCATTAATAAAGCTTGCAAGCTTAAT 474
Db 110 GlyAlaCysAlaIleProThrArgLysGluHisLeuLysSerHisAsnLeuGluSerIleGln 129
Qy 475 GATATTAAT-----ACCAAAATATTCATGATTTATGGAAATGACTAGTTACA 525
Db 130 ThrAsnAspLysArgSerSerLysAlaLysHisAsnLeuLeuAlaGluValCysTyra 149
Qy 526 GCAAAATAGCAAGGTGAATCAATTTGTAATTAATCAATCA-----CATAAAGCACT 576
Db 150 AlaLysPheGluGluGluSerIleValLysAsnTrpGluGluLeuGluLysHisThrThr 169
Qy 577 TCAGACGCTTGTACTGCTCTGTCAGCAAGCTTTGCAGATATGATATTTAGAGAGA 636
Db 170 GluGlyLysCysThrAlaLeuAlaArgSerPheAlaAspIleLysPheIleIleArgGly 189
Qy 637 ATNGATATGTTTAACCAATGTCATGAC-----AAAGTAGAAACGGGT 681
Db 190 LysAspLeuThrLeuGluLysAsnProGlnGluSerAlaArgTrpLysGluLeuGluAspAsn 209
Qy 682 CTCGAGAGGTTTCAAGAAAATACAT----- 708
Db 210 LeuArgLysLysPheGluLysLysLysLysGluLeuThrSerSerArgAsnGlyLysThr 229
Qy 709 GATGCAATGGAAGATGAAAGAAAATGATTAATCATCTGATGATCTGCAATTTATTAAT 768
Db 230 AsnGlyAlaGluGluLys-----AspLysSerGlyAsnTrpTrp 244
Qy 769 AAATTAAGAGACATGTTGGAATGTGAATGAATAAGATAGGAAGCTATTAACATGT 828
Db 245 LysLeuArgGluAspTrpAsnAlaAsnArgLeuAspIleThrLysAlaMetIleCys 264
Qy 829 GATGCAATCATATAATCTGATATTTATTAATCAATCAAGAAATATTAATCAATTTTCA 888
Db 265 LysAlaProGluLysAsnAlaProTrpPheArgAsnThrCysSerAsnGlyLysProThr 284
Qy 889 AATCTTAATATGCGCCATAAACAGCAAGAAAGTTCTACCAATTTAGATTTGTCCTCA 948
Db 285 GlyLulLysCysGlnCysLysIleAspGlyThrValProThrAsnLeuAspTrpValProGln 304
Qy 949 TATTTACGTTGGTTCGACGATGGGAGAAAGACTTTGCCGAAAAGAAATATTAATTTG 1008
Db 305 TyrLeuArgTrpPheGluGluTrpAlaGluGluLysPheCysArgLysAsnLeuLys 324
Qy 1009 AAAAGGTCAAGACCTCTGTCT-----AATGACAAAAGACGCTTATTTGATGT 1059
Db 325 GlnAsnAlaIleLysAsnCysArgGlyMetAspAspAspLysGluLysTrpCysSer 344
Qy 1060 CATATGACATGATTTGACACACATATTTGCAAAAAGATATTTGCAATTTGCAATTAAT 1119
Db 345 ArgAsnGlyTrpAspCysThrLysThrIleArgSerIleAspLysTrpSerMetAsnArg 364
Qy 1120 AAGTGTACTGACTGTTGCAATTAATGCAAAAGTTTGAAGTTGGTAGGAATCAACAA 1179
Db 365 GluCysThrLysCysLeuTrpValCysAspProTrpValLysTrpIleAspAsnLys 384
Qy 1180 GAAGCAATTTAAAAAAGAAAAAATATGAAGAAAGAAATACATATTTATTAAGAAC 1239
Db 385 LysGluPheGluLysGlnLysLysLysCysGlnAsnLulIleTrpArgAsnAsnGluSer 404
Qy 1240 GATTAACAAATTTGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1299
Db 405 SerGlnAsnSerProLysAsnTrpAsnAsnMetCysGluThrAspPheTrpGlyAsnLeu 424
Qy 1300 AAGCAAAACCAATATGCACTAATGACACTTTTAAATTTACTAATTAAGAAAGTAT 1359
Db 425 LysLysAspTrpGlnSerMetAsnAsp-----PheLeuLysLeuLeuAsnSerGluThrPro 443
Qy 1360 TGTAAAGAGAGATTACAGAGAAAAGATATTAATTTCTTAACAGTGTGTGATGACAA 1419

```

Db 444 CysThrAsnIleIleAspAlaLysSerLysIleAspPheThrLys-----AspProGlu 461
QY 1420 GGGATATTTTACGTCGATATTTCCCAAGTCGTCCCGACGCGGGGTCACAA----- 1473
Db 462 GluThrPheSerHisThrGluTyrCysAspProCysProTyrGlyLeuLysThrGln 481
QY 1474 TGTGATGATTAATTAATACACACACAAATCAGATTAATGCGTGAACGTGTAATATGAA 1533
Db 482 AlaAspGly---ThrTrpLysArgLeuTyrGluAsnAspProGlnCysProIleLysPro 500
QY 1534 GACATATTAACCTCCATGGGTCGATGACCTACATATATCATCTGCTTTATAGTGAAT 1593
Db 501 LysTyrGluProProTyrGlyValGluProThrGluThrAspValLeuTyrThrGlyLys 520
QY 1594 GAACAGCGTATATTCACAAATAATAGAAATTTTGTAAACAGCTCACTCAATATTCACAA 1653
Db 521 GluAsnLysAspIleIleValLysLeuArgGluPheCysLysThrAspGluAsnThrGly 540
QY 1654 GATTAATAATGCAAAATGGGAATGCTATATTAAGATGAAATATTAATAGATGTAA 1713
Db 541 PheLys---AsnGluGluTyrAsnCysTyrTyrGluValGlyLys---AspLysCysVal 558
QY 1714 CTGCAACAAATATGCAATCAATCAATATGATATCTAAGATATATCATTTCAATTTT 1773
Db 559 LeuGluAsnGluGluGluLeuGluGlyGluLys---LysValLysAspTyrAspAsnPhe 577
QY 1774 TTTCGATATATGCGTTACATTTTATTAAGGAGTACTATTAACTGAATGACAACTTAA 1833
Db 578 LeuMetPheThrValAlaHisMetLeuLysAspSerIleGluTyrPheSerLysLeuSer 597
QY 1834 ACTTGTATTAATATGCAACACACGATGATGATGATGATGATGATGATGATGATGAT 1893
Db 598 AsnCysLeuLysSerAspLysLysThrCysIleThrCysHisAspAsnGlyCysGlnCys 617
QY 1894 TTTCGACAGATGCGTTAAACAAAGAAGAAGATGGAATGATTAAGAATGCTGTACA 1953
Db 618 TyrAspLysTyrIleGluLysLysValHisThrGluIleLysLysHisPheAsp 637
QY 1954 AAAAAAATAATATACAG---CAATGCTATTAATGATGATGATGATGATGATGATGATGAT 2010
Db 638 LysGluThrAspPheGluIleTyrPheLysThrPheValLeuGluThrValLeuGluGly 657
QY 2011 ---TATTTTAAAGTATGATTAAGTGAACAAAGTGAACAAAGTGAACAAAGTGAAC 2067
Db 658 AspGluPhePheThrAspIleThrLysAlaTyrGluAspAlaArgGluIleValHisIle 677
QY 2068 ATGGAATAATATAAAGAAAAAATGCTTTTCCATTTTGGAAATTAATAGGACATAT 2127
Db 678 GlnGluMetLeuGlnLysLysGluGluVal-----LeuHisGluAspAlaSerAsn 695
QY 2128 TTAGAGAAATGCAATGACACTTTGATAGTCACTTAAAGAAAGAACTGCACAGATATG--- 2184
Db 696 MetLysThrIleIleAspGluLeuLeuAspHisGluLeuLysGluAlaLysGlnCysIle 715
QY 2185 -----AAGACAAATAT-----ACAAAGCAAGCATGTGAACA 2217
Db 716 ValAsnHisLysAspAsnAsnCysProAlaAspLeuSerAspSerGluAspGluGluGlu 735
QY 2218 TCCCATATGCAACAAACCCGTCGTAAACCTCTGAGAGCCAGCCAACTCAAA 2277
Db 736 AspIleProGlnArgGlnAsnLysCysAlaLysPro---SerGlyThrHisIleArgAla 754
QY 2278 AATTTAAAGAAATAGCAACAATCTTTAAAGAGATGCAATGCAAGCAAGCAAGCAAGT 2337
Db 755 LeuValAsnLysValAlaSerAsnMetHisLysLysLysArgGlnLeuValAsnArg 774
QY 2338 GGTGCT---CATTAATGCAAGAAAGCAACACAGATATATTAACGTGGGGGTAGG 2394
Db 775 GlyValSerLysLysLysGlyAspAlaLysGlyGluTyrArgLysSerGlyThr 794
QY 2395 AGAAGCACTTCAAGCAATTTATGTAATATGATTAATAACCTGTAATCGTAATCTT 2454
Db 795 ThrIleLysLeuLysAsp---IleCysSerIleThrAspAspHisSerAsnAlaLysArg 813

QY 2455 GGTTTTCAATGACCATGTGATGCAAAAGCAGCATGATGATATACAAACAGATTT 2514
Db 814 GlnHisThrAspIleProCysLysArgLysAspSerLysValAsnValLysAsnArg--- 832
QY 2515 GTGCTGGAACGTGAATGGGAAGTGGATCCGGAACACATCCGTAAGAACATCAGAGATGTT 2574
Db 833 -----ArgTrpMetAspThrAlaGlyPheIleSerAsnThrTyrLysAspIle 848
QY 2575 ATTTATGCTCTAGAAAGACATATATGATATGATATGATATGATATGATATGATATGAT 2634
Db 849 TyrMetProProArgArgGlnHisPheCysThrSerAsnLeuGluTyrLeuGlnThr 868
QY 2635 GATCAACCCCTTAAGTGAATATGTTGAT-----GATTAGTAAATATCTTT 2685
Db 869 AsnLysLeuLeuAsnGluLysAsnLysAsnLysAsnLysAsnLysAsnLysAsnLysAsn 888
QY 2686 TTGGGGATGCTCTTCTATCAGCAAAATATGAAAGCAACAAAGATTAATGATGATTA 2745
Db 889 LeuGluAspValLeuPheAlaAlaAsnTyrGluAlaAspPheIleLysLysMetTyrAsn 908
QY 2746 GAAAGAAATTAACCTTAAGGGCCCAAGAGATTAAGTGAACCAACACAGACATATG 2805
Db 909 LysGlnAsn-----AspTyrLysAspAsnAlaThrIle 919
QY 2806 TGTGAGCTATATGCTTACATTTTCCAGATATAGTGTATATATTCAGAGAGA----- 2859
Db 920 CysArgAlaMetLysTyrSerPheAlaAspLeuGlyAspIleIleGlnArgGlnHisIle 939
QY 2860 -----GATCTCTGGGAAGAAAGCGTGACATG 2886
Db 940 CysArgIleMetIleValGluArgValLysHisGluIleSerGluAlaGluAsnPheLeuIle 959
QY 2887 GTAAGCGTCAAGCAATTTGGAACCTGTTTGTGTAATATACATTAAGTCACTCAAGCGC 2946
Db 960 LeuSerLysLysAsnIleLeu-----AlaPheLysGluIle----- 971
QY 2947 AAGGAATGATTAATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3006
Db 972 -----TyrLysGluAspThrPro---TyrThrLysLeuArgGluAspThr 985
QY 3007 TGGGAGCTATTAAGCCCAAAAGTATGGAAGCCATGAATGTGATATTAATATTTGGAAG 3066
Db 986 TrpGluAlaAsnArgLysLysIleTrpGluAlaMetGlnCys----- 999
QY 3067 GATTAATCGGACACCAATACCAACAAAGTATGATGCGGATATGATGATGATGATGATGAT 3126
Db 1000 -----ProThrProAsnGlySerPheProCysLysSerTyrHisIleGly 1014
QY 3127 TTGATGATATATCCCAACAAATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 3186
Db 1015 LeuAspAspTyrIleProGlnArgGluArgTyrMetThrGluThrAlaGluTyrPheCys 1034
QY 3187 AAGGTGCAAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3246
Db 1035 LysGluGlnLysGlnTyrGluGluLeuValSerAlaSerAsnGluLysLysAsp--- 1053
QY 3247 GATTAATGCTCAAGCTGTACGAAGAGAGTACAGGTGTGCAAGTGAAGTGAAGTGAAGT 3301
Db 1054 GluArg-ValLysValValArgIleArgValHisAsnValGlnArgAlaCysLysHisVal 1073
QY 3302 -AAGCTGTATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3360
Db 1073 LysIleIleLysAsnLeuLeu---IleHisGly-----LysGluIleThrAspLysMet 1090
QY 3361 TCAATTAATATCAAGAAATTAATGATGAAAGCAACAAATGCTGTATGATGATGATGAT 3420
Db 1091 GluIleLysTyrLysLeuLeuTyrLeuGlnAlaGlnThrAlaAlaAsnGluIlePro 1110
QY 3421 GAAGCTCAAGTACTGCCAAATATATAGACAGCAAGATGATGATGATGATGATGATGAT 3480
Db 1111 AspThrTyrSerGluLeuValAspGluAsnGluLysProValValAsnPheLeuPheGlu 1130

Db	1578	-----	-----	AspLysValIlysLeuAspGluAsn	1588
QY	5302	TTGAGAGAAATATTTAACAAAATGACATCA	-----	-----	GTGGCGAAGGAAGT 5349
Db	1586	-----	:::	serGlyThrSerProLysIleValValProAlaProLys	1598
QY	5350	GATGACACTACAGGAATCCCGGTAGTACTGCGCGCAAAATTTTCTGTGAGACAAAATTAAG	-----	-----	5409
Db	1599	ProThrThrThrPheProProProSerProThrSerPhe	-----	-----	1612
QY	5410	GAATGTGTGTGAGACGAATGATATGCGGTACAAACCTGGTAGGAGATGAGAAATAGT	-----	-----	5469
Db	1612	-----	-----	-----	1612
QY	5470	GGAAATAGTGCAGAAAGATGAGAGATCTAAAAAANGTGGTCTGTAGCTTCAGATGAT	-----	-----	5529
Db	1613	-----	-----	-----	1615
QY	5530	GATTATTCCTATGGGGAAAAATCGCGATGAGAGTACTGGTATCGATTCTTCGATGGTTT	-----	-----	5589
Db	1616	-----	-----	-----	1621
QY	5590	GCCGATGGGGTGAAGATTTTTCGAACATTAAGAAAAAGAAATTTGGAATTTGTATGGG	-----	-----	5649
Db	1622	GIUGLUTPALAGLUTThrPheCysArgGluArgLysArgLeuGluLysIle	-----	-----	1639
QY	5650	GCGGTATGATTTTACTGTCGGATTAAGAACATTAAGAAAGAAATGTACA	-----	-----	5703
Db	1640	-----	-----	-----	1640
QY	5704	---GATCGCTACACAAATATAAAAATTTTATAGTAGTGAGAACCCAGATATGAAAAA	-----	-----	5760
Db	1657	GlyIuAspCysGluGluIleArgLys	-----	-----	1665
QY	5761	CAATCAAAAAATGCTGAGATTAAGAACAAAAATATATCCGAGATCCTGTGGCAAA	-----	-----	5820
Db	1666	-----	-----	-----	1682
QY	5821	GATGCAAGAGACGCTCGCCGAATTTTACACAAACAAATTAATAAAATTTGTGAAATATA	-----	-----	5880
Db	1682	yrCys	-----	-----	1692
QY	5881	AGTGAGATTGATATATATAGTATGAAAGAGTGTCACACAGCATTAATCACTGATGT	-----	-----	5940
Db	1692	-----	-----	-----	1692
QY	5941	AATAGTCAAAATATGCCCGCATTCATTAGACGATGAACCAAAAAGATTGAAGAAAGTGT	-----	-----	6000
Db	1692	-----	-----	-----	1692
QY	6001	AATGTCAAGTGCCAGAGGTCCACACGCTGTACGAMGGGAACACCGTACACCGGTA	-----	-----	6060
Db	1692	-----	-----	-----	1692
QY	6061	TCACGTATTCAAAAAGCCGCGCATCGAAAAAAGAGCGAAAAACGCGCTTCACAAA	-----	-----	6120
Db	1693	-----	-----	-----	1706
QY	6121	CAGCGAAAAAAGTGGAAATCTAATCAACAGAAAAATGCGAGACAAACGAAACCGCAGCA	-----	-----	6180
Db	1706	snGlnLysThrAspAlaArgGlnAsn	-----	-----	1721
QY	6181	GCAGCACAACAACAAGAAACGACATCAACAGCACAACACAGATTCGAGTGGGC	-----	-----	6240
Db	1721	erThrThrLeuAspHrcysThr-ThrAla	-----	-----	1731
QY	6241	ACAATGTAAAGCCATCTTTCGATTAACACAGATACGAGGTTGAATAGAGGTTGT	-----	-----	6300
Db	1732	AspPheLeuGlnThrLeu	-----	-----	1737
QY	6301	AATCCAAAACGTATGACAAATATCTAAATGGGTGTATTTAGTAGTAAAGAA	-----	-----	6360
Db	1737	-----	-----	-----	1737

QY	6361	ATGAAATGGCATATGCTATGCTCTCTAGAGAAAAAAATTTATGTAATAATATACAA	1742
Db	1738	---: --- ---LysAsnGlyProCys---	1742
QY	6421	TATTTAAATTATGAACGTAAATTAAGCGTGCAATGATATATAAGCGCTTTATATAA	6486
Db	1743	---: --- ---LysAsnAspAsn---	1748
QY	6481	TGTGCAGCAATAGAAACTCAATTTTTGTGTAAATATATATTAATGAATAATCTGCACGA	6540
Db	1746	---	1746
QY	6541	GAAATGATTCGCAAAATGGAACAATTCAGATGCAATTTAAAGCAATATGATATATACA	6600
Db	1746	---	1746
QY	6601	TATGTGATTTAAAGATATGTTTTTTGGAACGTATATTTCTTAATGATATAAAATATATA	6660
Db	1747	---: --- ---ValAspSerGlyGluAsnIys---	1754
QY	6661	ACTGTAACAATATGTTACACAAACATTTCTCAATGAAATTAATAAGAAAAACAGATATA	6720
Db	1754	---	1754
QY	6721	AAAAAGATGAAGATTAACGTAAATATTTTTGGGAGAAAAATATAAAATTTATTTGGAA	6780
Db	1755	---: --- ---LysIlePhe---	1757
QY	6781	GGATTGATATATGATTAACCTATATCATCTCACAGACAGCAAAACGAAAAAATTAGA	6840
Db	1758	---: --- ---AspGluAsn---	1761
QY	6841	GATATTTACCAAGTACATGACATGACCAAACTGACGCTTCCTTGAGAGTTGTATAA	6900
Db	1762	---: --- ---AspThrPheIysTyrThr---	1767
QY	6901	AGGCCCCAATTTTGGAGATGCTTTCACAGAAATGGGACAGAAATTTTGTATATAGAGAG	6960
Db	1768	---: --- ---GlnTyrCysGly---	1771
QY	6961	GAACAGTTGTTAAATTTGGAGCGGCGCTGTAAAGATATAGAGTATATGTTACTATATAC	7020
Db	1772	---: --- ---ThrCysSerIleAsnGlyPheIysCysAsnIlyAspSp---	1784
QY	7021	GGTAACACACAGAAATGTGCAGAGCGCTGTATACATATCAAAATTTTATTAAGAAGTGG	7080
Db	1785	---: --- ---CysArgValArgThrAsn---	1790
QY	7081	AAAACGTATATGAAAGACAAAGAGAAAAGTTCAAAAAGATTAAGATGCGCAAAAGTAT	7140
Db	1790	---	1790
QY	7141	AAGGATTAATCCTTCTACTGAAAAGACATAGAGAAGCAACATGCTGCTCATGAAATATTTA	7200
Db	1791	---: --- ---ValThrCys---	1793
QY	7201	AACATGAATTAAGAAATATATGTGCAATAAGATTTGTTCTGTATGCAAAAACCTTCT	7260
Db	1793	---	1793
QY	7261	TCACAATTAACAAAAACAACAACAATCAACATCCGATGCTTAATGATATGCGAA	7320
Db	1794	---: --- ---AsnGlySerAsnArgThrThrIleThrAlaAspIle---	1807
QY	7321	TGCGTGATTAATGTTCCCTGGAAGATTTTAACAAGTGTGAGTGTCTGAACTTCAAAAAAG	7380
Db	1808	---: --- ---LysAsn---	1809
QY	7381	GGATCTATGATTCATACAAAAAAATTTACTGACCTAAATATCTATGATATGTGTAGAG	7440
Db	1810	---: --- ---Gly---	1810

QY 7441 AAGCAGCATTTATTTATCTTAAGACGAGAAATATATG-----GATATT 7488
DB 1811 -----GlySerSerAlaGluLeuMetLeuValSerAspPhe 1824
QY 7489 ACCCTGAGCAAAAAATTTTACCTTAGTGTACAAAGGAAAAAGAAAGTAAAAATAG 7548
DB 1825 AsnSerGlyAsnGlyPheAsnAspLeuGluAla ----- 1835
QY 7549 TGGACTAATTAATTAATCCTTGGATCCTTAAGAAACCTTATGCACCTGATTAATATATAGGA 7608
DB 1835 ----- 1835
QY 7609 AGAAGAAACCCCTTGTGAAAATAGAGAAATGCTTTTAA-----GTAGATTATGAA 7662
DB 1836 -----CysLysAsn-----AlaAsnIlePheGlyGlyLeuSerGluAsnLys 1849
QY 7663 TGGAAATGTTCAAAATTTCAAGTCTTATCAGAGAGAAAAAGAGTATGTACCTCCA 7722
DB 1850 TTPlyScys-----ValTyrPheCysLysSerAspValCys----- 1861
QY 7723 AGAAGAGAACATATGTCTTAAGGAATTTAGATGAATTTAAATGAAAGACTTAAGAT 7782
DB 1861 ----- 1861
QY 7783 AGTAAATATCTCTTAAAAATGTTGCTGCAACTGCACGAAATGAGGAATAGCATATA 7842
DB 1861 ----- 1861
QY 7843 AAAAATCTCACTCAGAGAACGGGTGCGCAATGATCAATATGTATCTATGAATAT 7902
DB 1861 ----- 1861
QY 7903 AGTTTCGTATCTGGGTGACATATGTAGAGAACAGATATGTTACGAATTTGCTGTAC 7962
DB 1861 ----- 1861
QY 7963 TTAACCTCCGTAGAAATTAATATTAAGTTTGAATACATATATGAAAAATGAGA 8022
DB 1861 ----- 1861
QY 8023 AATTAATAATTAAGTAAATTAATCAACAGATGTACAAAAGTTGCTGTCTGTGTG 8082
DB 1862 -----GlyLeuLysLysAsnAsnAspPhe----- 1869
QY 8083 GATGCTAATAGAAAGATTTTGAAGAAGCAATGCTGCACCAAGCAGAAAGATGCAAAA 8142
DB 1870 AspGlnAsnGlnIleLeuLeuIleArgAlaLeu----- 1880
QY 8143 CTTTTTGAAGAAAGAAATGATGATGATTTGAACGATTAATTAACAAGATAAGTGT 8202
DB 1881 -----PheLysArg----- 1883
QY 8203 GGACATAGAGAGATCCACCTGTGATGATTAATATCTCAACGTTTCGATGATGAC 8262
DB 1883 ----- 1883
QY 8263 GAATGCTGAATATTTATTTAAGCAGTGAAGAAATTTGAAAAATTTAAAAATCA 8322
DB 1884 ---TTPLeuGluTyrPhe-----LeuAspAspTyrAsnLysIleArgLysLys 1898
QY 8323 TGTGATCATGTAAACATCTGACAGATGCAAGATGATTAATGATTAATAAGTGTGA 8382
DB 1899 LeuAsnProCys-----IleAsnAsnGlyGluLysAla 1909
QY 8383 CAGGTGAACAGATGTCAAGATTAATTAATTTTGTCTTAATGAGAAATCTATATTC 8442
DB 1910 TleCysThrAsnGlyCysValGlu-----GlnTTP----- 1919
QY 8443 GATATACAAATCAAAATTAATCAAAAGATTTGTATGAACAACCAATATATCAAAAAATCTCT 8502
DB 1919 ----- 1919
QY 8503 ACTTATGATCATGTCAAAATTTTGTACAAAAGTGAACCTTTTAAAGTGAATGTCT 8562

DB 1920 ---IleAsnHisLysArgTyrGluTyrPheAsnLeuLysSerPhe----- 1933
QY 8563 GTTGAGAGCTTTCTGAAATATCTTCAATGAGAAACAAGTAAAGCTTTGAATTAATTAAT 8622
DB 1934 ----- 1934
QY 8623 GAAATATGATGCTTCTTCAATATACGAAACATATGCTTTCGAAACACCAAAAGTTAT 8682
DB 1939 GlyAspAsp----- 1941
QY 8683 AAGAAAGCTTGACGTGTACACTACTCTTAAGATTCATGATTAATTTCTCCACGAT 8742
DB 1942 ----- 1942
QY 8743 CAAGCAAGATGATGATTAAGAAATTAACAACCTTACCTTCTGTCGAGAAATGATAT 8802
DB 1944 ArgAsn----- 1945
QY 8803 GATTAATTAATCTGTAATTAATGGAACGATACCTTGTCTTAATAGTTCAGATGATACAA 8862
DB 1945 ----- 1945
QY 8863 GGTGTATGATTTCTCCAGAGAACAGACATTTATGTACAAAGACCTATCACTGCATTAAT 8922
DB 1946 -----ProArgLeuArgPhe----- 1950
QY 8923 TATAGAAAAGTGATTAAGAAATTTTAAAAAAAACCTTACTTCTGCTTCAGTCAA 8982
DB 1950 ----- 1950
QY 8983 GGACAAATTTGTAGTCAAAAATTAATTAATCGAAGAGAGTTGTGCTTGAAGCAATGAAA 9042
DB 1950 ----- 1950
QY 9043 TATAGTTATGACATATTTCCGATATATTAAGAACTGATATGATGCACCTTCAATTA 9102
DB 1951 -----PheValAspLeuIle----- 1955
QY 9103 TCTGAAAAAATTAATAATTTGAACATCAATGAACGAAACGAAATTCGTAAGAA 9162
DB 1955 ----- 1955
QY 9163 TGGTGGAAAAATTAATAGCTCAGATATGCGACGCTATGTTATGAGATTAATAATTTGCT 9222
DB 1956 -----ArgGlnIle----- 1958
QY 9223 ACTTCAAAAGTAACTATTAATTAAGTGAAGTGTGTCAATTAACCAAGATGAAGAACTAAT 9282
DB 1959 -----AlaAlaThrIleAspLysGly-----Asn 1966
QY 9283 CAGTTTCTGCTGTGTTAATTTGAAGGCAAGCAAGCATGTAAGGAAAGAAACATGTA 9342
DB 1967 His-----AsnGlyLeuValLysLeuValLys----- 1975
QY 9343 AGTATTCATTAATAACAAATGCTCCTGTTCAAGCAAGATATTTTGAAGCGTAGAA 9402
DB 1976 ---SerValLysCysAsnCysGlyAsnAsnSerGlnAsnGlyLysGluGlu 1993
QY 9403 TTAATTAAGACACCTGATGTCAGAATGATTAAGAAATATATTTAGCTTGAATATATG 9462
DB 1994 -----AsnAspLeu 1996
QY 9463 ATAAAAAATACATGAGAAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9516
DB 1997 ValLeuCysLeuLeuGlnLysLeuGluLysLysAlaGluLysCysLysAspAsnProGlu 2016
QY 9517 TCTTCAGGTAAATTAAGACATTAACCATCTGAAGAAATGTTAGCTATATATTAATTAATCA 9576
DB 2017 ThrSerGly---IleProGlnGlnProCysGluValSer---ProAsnHisIleGluAsp 2034
QY 9577 AAGATTCATCAATGCGCTTGGAGTTAATGATAT-----AATGAA 9618

```

QY 10546 AGAATGATTCTTATAGAGGTATAGATTAAATTAAATGATTCAC-T-AGTGTGCTAAACCT 10604
Db 2387 HISANSNVALTYRISerGIleaspleuileasAspThrIleSerclLysnclunHis 2406
QY 10605 ATGTATATATATGATGAAGTATGG 10628
Db 2407 ILeaspIeTyrisAspGluleuLeu 2414

RESULT 14
Q26033 PRELIMINARY; PRT: 2664 AA.
ID Q26033;
AC Q26033;
DT 01-NOV-1996 (TrEMBLrel_01, Created)
DT 01-NOV-1996 (TrEMBLrel_01, Last sequence update)
DE 01-DIC-2001 (TrEMBLrel_19, last annotation update)
DN Variant-specific surface protein.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RL infected erythrocytes."
DR EMBL: L40609; AAB75398.1; -.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PEEMP.
SQ SEQUENCE 2664 AA; 302410 MW; 6EA2468511703091 CRC64;

Alignment Scores:
Pred. No.: 1,29e-129 Length: 2664
Score: 2726.50 Matches: 906
Percent Similarity: 34.34% Conservative: 333
Best Local Similarity: 23.95% Mismatches: 879
Query Match: 14.01% Indels: 1605
DB: Gaps: 122

US-10-087-013-1 (1-10628) x Q26033 (1-2664)
QY 97 AAAATGCAAAAATGTTTGGAAGCTTAAGCCAAAAATTAAAGACATCATCAAAATPAT 156
Db 17 ArgSerAlaValSHisLeuAspSerIleGlyLysIleVal-----Tyr 31
QY 157 GCAGAAGAACATGCGANTCG-----TTGAAAGGCATTTGACG 195
Db 32 AspIysValHisGLyAspAlaIleuGlnProSerAsnGlyLysLeuIysclYthrIleuSer 51
QY 196 AAACGACAATTTGCTGGTGTCTCTTAGCCGAGTAATAAGCATAATATATATATCA 255
Db 52 LeuAlaIlePheGluIuysAlaProGluGlyLysGlnThrSerGlu----- 67
QY 256 TATCCATGTAATTTAGATCATTAAGAAACATTAATTAATTCAG-----TATGATGANTGTG 309
Db 68 ---ProcysAspIeuAsnhIsGluIuYrHISThrThValThrSerGlyTyraSplySGlu 86
QY 310 AATTGAGACATCTTGGCCATGGTAGAGACAAACCATTGTGATGACAGATGAAGAAACT 369
Db 87 Asn-----ProcysIysAsparArgProGluValAlaArgPheSerTyrThrGluGlyAla 103
QY 370 GAATGTGGA---AATTAATACGTAATTTTAAAAAAGAAAAATGATGCTAGTACCTGTGCG 426
Db 104 GlucIysAspIysSerIuysIleArgGlySerAsnSerAsnIusAspGly---AlacysAla 122
QY 427 CCACTAGAGAAGACGATATGTGTATATAAAACCTGGAAGCTGTA---AATGATATAAAT 483
Db 123 ProPhaArgArGleuHisIeuCysAspGlnHisStrGluHisIleTyrGluValTyrGluVal 500

```

Alignment Scores:	
Pred. No.:	1,29e-129
Score:	2726.50
Percent Similarity:	34.34%
Best Local Similarity:	23.95%
Query Match:	14.01%
DB:	5
Length:	2664
Matches:	906
Conservative:	393
Mismatches:	879
Indels:	1605
Gaps:	122

QY 4675 ATTGCTAATGGTTACAGACTATCTAGAGCAAGATGCAACAGATTACTTGAACAGAAA 4734
 DB 1262 ----- 1262
 QY 4735 TTACTGCTAGTTGGTGTATAAGCTGAGAGTGCCTGTGTGTACAAAGAAATATACAA 4794
 DB 1262 ----- 1262
 QY 4795 TTGTAGAAAACAGCGCTTACTATGATGCGCAACAACTTTGGTGCACAAATTTATTT 4854
 DB 1262 ----- 1262
 QY 4855 GAAATGACGACAAATATCTAACAATTTGAGTAAAGATAGTGAAGAGATTAGTAAAG 4914
 DB 1262 ----- 1262
 QY 4915 GAGGCAACACAGTGCCTTAAAGTGCACAAACAAAGTCTTAATTAATCAATTAATCTTG 4974
 DB 1263 ----- 1273
 QY 4975 AAGAATTAAGTGAAGATGCTTTCTCTCTGCTGACATGATATGTTTCATGCA 5034
 DB 1274 ----- 1278
 QY 5035 TTGGATGCAATTTATACAGATCCAGAAATTAAAGATGAAATGGGTTGCGAAAAAGATTG 5094
 DB 1278 ----- 1278
 QY 5095 ATGGAATGCGCGCAACGAGAGGTACATTTGGTCACTACTACAAAGAAAAAGAA 5154
 DB 1278 ----- 1278
 QY 5155 AAGAAGAAATATAAACGTCGATGCGCACAATATTCTTATGAGTCCGCTGTAGT 5214
 DB 1279 ----- 1283
 QY 5215 GCTATGAATATAGTTTATGATTTAAGATATATATCTAGTATGATTAATTTGAA 5274
 DB 1284 ----- 1296
 QY 5275 GATGAAAAACAAAGCAGAGGAAATTTGAAGAAATTTTACAAAAATGACATCA 5334
 DB 1296 ----- 1296
 QY 5335 GTTGCAAGAGATGATAGTACTACAGAAATCCCGTAGTACTGCGGAAATTTTTC 5394
 DB 1296 ----- 1296
 QY 5395 TGGACGAAATTAAGAAATGTGTGTGAACGCAATGATGCGGGTACAAACGTGTAG 5454
 DB 1296 ----- 1296
 QY 5455 GATGATGAAATAGTGAATAGTGCAGAGATGATGAATTAATAAATTTGTTCT 5514
 DB 1297 ----- 1301
 QY 5515 GTACCTTCAGATGATGATTTCTCTATGAGGAAATCCGATGAGAGTACTGCGTATCAG 5574
 DB 1302 ----- 1308
 QY 5575 TTCTCTCGATGTTTCCGAAATGGGTGAAGATTTTTCGAAACATTAAGAAAAAGAAATG 5634
 DB 1308 ----- 1308
 QY 5635 GAGAAATGTAAGGGCGGTATATATTAATCTGTGTGATTAAGATTAAGAAAG 5694
 DB 1308 ----- 1308
 QY 5695 AATATGACAGATGCGTGTACACAAATTAATAATTTATTAGTACGTGAACACAGATAT 5754
 DB 1308 ----- 1308
 QY 5755 GAAAAACAAATCAAAATATGTTGAGAAATTAAGACAAATATATTTCCGACATCTGTG 5814
 DB 1309 ----- 1310
 QY 5815 GCAAAAGATGACAGAGACCGCTCGCAATATTTAGACAAACAAATTAATAAATTTGTGAA 5874
 DB 1310 ----- 1310
 QY 5875 AATTAAGTGAAGATTTGATTAATTAAGTATGAAGATGTCACACAGCATTAAT 5934
 DB 1311 ----- 1323
 QY 5935 GATGTAATAGTCAAAATATGCCCCATTTAGACATGACAAAGAGTTGAAG 5994
 DB 1324 ----- 1328
 QY 5995 AAGTGAATTTGCAAGTGCACAGAGTCCACAGCTGACAGAAACACCGTCACCA 6054
 DB 1329 ----- 1339
 QY 6055 CCGGTATCACTGATATCAAAAGCGACGATCGAAAAAGAAACAGCGCGCT 6114
 DB 1340 ----- 1354
 QY 6115 ACAAAACAGCCGAAAAAGTGAATCTAACACAGAAATGGAGCACAAACAGAAC 6174
 DB 1355 ----- 1364
 QY 6175 CGACGACGACACAAACACGAAACGACATCAACACACACAGATGAC 6234
 DB 1365 ----- 1380
 QY 6235 GTGGCACAATGTTAAAGCCATTTCTGCAATTAACACAGATGACAGGGTGAATAGAC 6294
 DB 1381 ----- 1387
 QY 6295 GGTGTAATCCAAAAAGTATGACATATCTAAATGGGTTGATTTAGTAACTCT 6354
 DB 1388 ----- 1395
 QY 6355 AAGAAGAAATGAATGCAATATGATGCTCCCTCAGAGAAATTAATCTATTAATAT 6414
 DB 1396 ----- 1401
 QY 6415 ATCAATATTTAATTAATGAACTGAAATTAACGTCACATGATTAATAAGAGCTTTT 6474
 DB 1402 ----- 1414
 QY 6475 ATTAATGTCACCAATAGAACTCAATTTTGTGTTAAATATTAATGAA 6528
 DB 1415 ----- 1434
 QY 6528 ----- 6528
 QY 1435 ATAGLILysLysLaclLueuGlnasnGlyLLeuaspLeuTyrSerSerGlyaspGly 1454
 DB 6529 ----- 6579
 QY 1455 AspProaspasnPro-----GlnasnLysLeuLeuasnGlyValIleProProasphe 1472
 DB 6580 ----- 6639
 QY 1473 LeuArgLeuMetPheTyrThrLeuGlnAspTyrArgAspIleLeuValHisGlyLysn 1492
 DB 6640 ----- 6699
 QY 1493 ThrSerAspSerGlysnThrAsnGlySerAsnAsnAsnIleValLeuGlnValSer 1512
 DB 6700 ----- 6747
 QY 1513 GlyAsnLysGlnLysPmetGlnLysIleGlnGlnLysIleGlnLysIleLeuProLysAsn 1532
 DB 6748 ----- 6759

Db	1533	GLYGLYThrProLeuValProLysSerSerAlaGlnThrProAspLysTrpTrpAsnGlu	1552
QY	6760	AAATAAAATTTATTTGGGAAGGATGATATATGATTAATTAAT	6804
Db	1553	HisIaGlnSerIleThrPylsGlyMetIleCysAlaLeuThrTrpGluLysAsnPro	1572
QY	6805	-----CATTCACAGCAAGAAAGAAAGAAAAAATTTAGATTAATTAAC-----	6849
Db	1573	AspThrSerAlaArgLysAspGluAsnLysIleGluLysAspAspGluValTyrGluLys	1592
QY	6849	-----	6849
Db	1593	PheheGlySerThrAlaAspLysHisGlyThrAlaSerThrProThrGlyThrTyrLys	1612
QY	6850	---CAGTACAAT---GACATGACCACAACTG-----	6873
Db	1613	ThrGlnTyrAspTrpLysValLysValLysLeuLysAspThrSerGlyAlaLysThrProSer	1632
QY	6874	-----ACGCTTCCTTGAAGAGCTTTTAAAGAGCCCAATTTTGGATGAG	6921
Db	1633	AlaSerSerAspThrProLeuSerAspPheValLeuArgProProTyrPheArgTyr	1652
QY	6922	TTTCACAAATGGCAGAGAATTTTGTAAATAGAGAAAGAGAGCTTTTAAATTTGGAG	6991
Db	1653	LeuGluGlnTrpGlyGlnAsnPheCysLysGluArgLysLysArgLeuLysGlnIleLys	1672
QY	6982	GGCGGCTGT-----AGGAATATGAGTGAATGCT-----	7011
Db	1673	GluGluCysMetAspLysSerAspLysLysTyrSerGlyLysGluGlnCysAspArg	1692
QY	7012	-----AGTAATGACGGTAAACAGACAAAGATGTGACAGCGCTGT	7050
Db	1693	ArgAspThrSerAsnGlnValSerAlaAspLeuGluLysArgSerCysGlyAsnSerCys	1712
QY	7051	GTACATATTCAAAAATTTTATTAAGAAGTGAAGAACTGAATATGAAGACAAAGAAAG	7110
Db	1713	ArgPheTyrLysLysTrpLysArgLysArgLysArgLysGluLysArgLysGlnIleAsnAla	1732
QY	7111	TTTCAAAAAGATTAAGATGAGCAAAAGTATTAAGATTAATTCCTTCTACTGAAGACATA	7170
Db	1733	TyrSerLysGlnLys-----ThrLysTyrGlu-----	1741
QY	7171	GAGAAAGCAACATGTCCTCATGAAATTTAAACATGAATTAAGAATTAATGTGCAAT	7230
Db	1741	-----	1741
QY	7231	AAGGATTTGTTTGTATGCAAAAACCTTCTTACAACTACCAAAAACAACACAATGCA	7290
Db	1742	-----GluGlySer-----	1744
QY	7291	CATATCCGATGCTAATGATATGCCAGATGCTGATTAATGTTCTGTAAGAAATTAAC	7350
Db	1745	LysGlyAlaGlyLeuAsnAsp-----HisAsn-----	1753
QY	7351	AAGTGTGAGTCCCGACACTTTCAAAAGAAAGGATCATGATTCATCAAAAAAATTAAT	7410
Db	1754	LysGluPheCysValLysLeuLys-----	1761
QY	7411	GAACCTAAATTAATCTATGATTAATGTAGAGAAAGCAGCATTAAT	7455
Db	1762	-----ThrCysThrAspAlaIleAlaPheLeuAsnArgLeuLysAsn	1775
QY	7456	-----TTATCTAAAGACGAAAAAT---AATATGATTAATTAACCTTG-----AAG	7497
Db	1776	GlyProCysLysLysAspAsnGlnLysAsnGlyLysAsnAspIleAsnPheGlyAsnThrGlu	1795
QY	7498	GAATAATTTATACCATGATGCTACCAAGAAAGAAAGAAAGTAAATATGTTGACTAAT	7557
Db	1796	GluThrPheArgProAlaGlnAsnCysLys-----	1805
QY	7558	AATAATCTTGCGATCTCAAGAAACCTTAATGACCTGATTAATATATATGGAAGAAAGAC	7617
Db	1806	-----ProCysSerSerPheLys-----IleAsnCysArgAsn	1810
QY	7618	---CCTTGTGAATAATAGAGAAAGAAATGCTTTTAAGTAGATTAATGAATGAATGATAC	7674
Db	1817	GlyAsnCysArgSerGlyAspGly-----AspThrLysGluLysCys-----	1830
QY	7675	AAAAATTAAGTCTATACAGAGAGAAAAAGAGATATGTACCTCCCAAGAGAGACAT	7734
Db	1830	-----	1830
QY	7735	ATGTCTTAAGGAATTTAGATGAATTAATTAATGAAGACCTTAAGATGATTAATTCCTC	7794
Db	1830	-----	1830
QY	7795	CTAAATATGTTGCTGCGAATCTGACAGAAATAGAGAAATGAATGAATGAATGAAT	7854
Db	1831	-----	1841
QY	7855	TCAGAGAACGGGTGCGCAATGATACCAATATGTGATCTATGAATATAGTTGGCTGAT	7914
Db	1842	ThrMetGlyThrCysThr-----	1847
QY	7915	CTGGTGACATAGTTTAAGAGAACAGATATGTTACGAATGGTGGTTACTTACTCCCGTA	7974
Db	1848	---GluAspValVal-----	1851
QY	7975	GAATAATTAATTAAGTTTTGAATACATATATGAAATATGAGAAATTAATTAATTA	8034
Db	1852	-----MetHisValSerAspLysAsnAlaAsnGlu	1861
QY	8035	GGTAGAAATTAATTAACAGATGTAACAACGTTCTGCTGCTGGTGAGATGTAATAGA	8094
Db	1862	PheGluLysAspLysLeuAspArgLysAlaCysGluAsnAlaGlyIlePheThrGlyIleArg	1881
QY	8095	AAAGATTTTGGAAAGCAATAGAGCTGGAAGACACAGAGATGCAAACTTTTGAAGAA	8154
Db	1882	LysAspLysLysLys-----CysArgLysValCysGlyLeuHisIleCysLysGln	1898
QY	8155	GGAAGATGAGATGATTT-----GAACGATTAATTAATTAACAAAGTATGATGGA	8205
Db	1899	GluLysGlyAsnGlyAlaIleAsnAspGlnGlnIleLeuValArg-----	1914
QY	8206	CATTAAGAGCATCACCCTGTGATTAATTAATCTCAACGTTTGCATGATGACTGAA	8265
Db	1915	-----AlaLeuLysArg	1919
QY	8266	TGGTGAATATTAATTTGTAAGACATGATGAGAGATTTGGAATTAATTAATTAATCATG	8325
Db	1920	TrpValGluTyrPhe-----LeuGluAspTyrLysLysIleLysLysLysLeu	1935
QY	8326	GATCAGCTGAATAACATCTACAGATGCAAGATTAATTAATTAATTAATTAATTAAT	8385
Db	1936	LysProCys-----IleGluAsnGlyAsnGlySerThr	1946
QY	8386	TGTAAACAGAGATGTCAGAAATATTAATAATTTTGTCTTAATTAATTAATTAATTAAT	8445
Db	1947	CysIleAsnGlyCysAsnLysLysCysAsnArgValGlyLysLysLysLysLys	1966
QY	8446	ATACATCAATTAATTAATTAACAAAGATTAAT	8475
Db	1967	AspGluTrpThrLysIleLysAsnHisTyrLeuGluLysAsnLysGluLysAspLysAsn	1986
QY	8476	-----GACACCAATATATACAAAATCTCTATGATGAT	8511
Db	1987		

QY 8623 GAAATGATGCTTCTTCATATAGACATATGCTTGGAGAAACACCAAAAGTTAT 8682
 Db 2031 LysAsnAspLysAsnGluAsn 2037
 QY 8683 AAGAAGCTTGAGTTGATGACCTTCTGAAGATCCATGATTAATGTCTACCGAT 8742
 Db 2038 ---AspAlaLeuAspCysHeu 2044
 QY 8743 CAATAAAGATGATGATGAAGATTAACAACCTTCTGCTGCAAGATGATAT 8802
 Db 2045 ---AsnLysLeuGluThrLysIleHisGluCysLysThrGlnHis 2058
 QY 8803 GATATAATCTGTATTAATTGGAACGATACCTTGTCTTAATGCTCAGATGAT 8856
 Db 2059 GluAsnSerValGluAsn ---SerAspGlnProHis 2069
 QY 8857 ---ACAAAGGCTGATGATTCCTCCAGAGAAAGACATTTATGTACAAAGACCTATACCT 8913
 Db 2070 ProAsnCysGlyGlyAsnProProPro 2078
 QY 8914 GCATATAATTATAGAAAGGTGATAAAGAAATTTTAAAAAAAGCTTCTACTCTGCT 8973
 Db 2078 --- 2078
 QY 8974 TTCAGTCAAGGACAAATTTGATGCTCAAAAATATTAATCGAGAGAGAGTTGCTTTGAG 9033
 Db 2079 ---AspLysGluLysLeuLeuGlu 2086
 QY 9034 GCATGAATATAGTATGATGAGATTAATCCGATATAATTAAGAACTGATATGATGAGAC 9093
 Db 2087 Gln 2087
 QY 9094 ACTTCATATCTGAAAAAATTAATAATATTTGAAACATCAATGACAGCAACGAAAT 9153
 Db 2087 --- 2087
 QY 9154 CGTAAACATGCTGGAAAAATATATAGACGTCACATATGCGACCTATGTTATGCTAT 9213
 Db 2088 ---GluAsnProValGluGln 2093
 QY 9214 AAATATGCTACTCAAAAGTAACATTAGATGAGATGCTCAATTAACCAAGATGAA 9273
 Db 2094 ---ProGlyPheCysProThrProGlnGlnGln 2103
 QY 9274 GAAACATATCAGTTCTGCTGCTGTTAATTGAATGGGCAAGCAGCATGTAAGAAAG 9333
 Db 2104 ProGluProAsp 2107
 QY 9334 AAACATGTAAGTATTCATTAAAAACAAATGCTGCTCAAAACGAGATTAATTTGAA 9393
 Db 2108 ---AspLysCysGlyLysLeuGluLysLysAspGln 2119
 QY 9394 GCGTCAGAAATTATTAAGACAACCTGATGTCACATGATTTGAAATATATTAGCTTG 9453
 Db 2120 LysLysGln ---GlnPro 2124
 QY 9454 AATATATTGATAAAATAACATGAGAAATCTAATATAAATAATAGCAATTAAAAAGAT 9513
 Db 2124 --- 2124
 QY 9514 CAATCTTCAGTATATAGACAATAAACATCTGAGAAATGCTCAGTATATATAAA 9573
 Db 2125 ---GluGlnProAlaGluGlnLysProGlyAlaIleValPro 2137
 QY 9574 TCAAAAGATTCATAGCGCTTGAGGTTAATGATATAATAGTAATAGTTACAGGACA 9633
 Db 2138 SerGlyPro 2140
 QY 9634 AAAAAATGAAAAATATGATTCAAAGAGTACTAAAAAATTAATATCCCTGTTTATAT 9693
 Db 2141 ---ProGly 2142

QY 9694 TTTGTTGAGATGAAACACACAAAAATGATGATGATGAAATATATAAGAAAGAG 9753
 Db 2143 ---SerGluProGluAlaAspLysGlyProValLysProAlaGlu 2157
 QY 9754 CAACAGTTGCTGCTAAGACAGCTATTTCTTACACCCCATGATGATTTCTTATCAA 9813
 Db 2158 ---ProLysPro 2161
 QY 9814 GCACCTTTATTTCAACACATGAGTACGACAAATGATGCTTAAATGATATATGAAA 9873
 Db 2162 GluProProAspLeuSerHisProAla ---ValIlePro 2173
 QY 9874 AGTAGATCTCTGTTGTTATGATGCGCTTACGTTG 9921
 Db 2174 SerLeuValThrSerThrLeuAlaThrSerValGlyLeuGlyPheAlaIlePheThr 2193
 QY 9922 CATTTCAAGAAAGAAAAATTCAAATGCTGCTG ---GACTGTTGCTGATACGAAATAC 9978
 Db 2194 PheTyrLeuLysLysLysThrLysSerSerValGlyAsnLeuPheGlnIleLeuGlnIle 2213
 QY 9979 CCGCAAGGAGATGATGAAATGCTACCTTGAATCCAAATAGATGATACCATATAGA 10038
 Db 2214 ProLysSerAspTyrAspIleProThrLysLeuSerProAsnArgTyrIleProTyrThr 2233
 QY 10039 AGTGCTCATATTAAGGCAAAACATATATATATGAGAGAGATACCT ---AGT 10089
 Db 2234 SerGlyLysTyrArgGlyLysArgTyrIleTyrLeuGlnLysAspSerGlyThrAspSer 2253
 QY 10090 GGAGATCAAGATATAATATGATGAGACTTATCTCCGCTATATTAATGATCGAACT 10149
 Db 2254 GlyTyrThrAspHisTyr ---SerAspIleThrSerSerGluSer 2267
 QY 10150 GAGTATGAGAAATGATGATTAATGATATATATGATACAGAGTGTCTTAATTAACA 10209
 Db 2268 GluTyrGluLysLeuLysLeuAspIleTyrValProLysSerProLysTyrThr 2287
 QY 10210 TTGATAGAACTAGTACGACACATCAAAAGGATATACCAAGTAT 10257
 Db 2288 LeuIleGluValIleGluGluProSerGlyAsnAsnThrThrAlaSerGlyAsnAsnThr 2307
 QY 10258 ---GATACACCAACT 10269
 Db 2308 ThrAlaSerGlyAsnAsnThrThrAlaSerGlyLysAsnThrProSerAspThrGlnAsn 2327
 QY 10270 ---AATGATACACAGCAGTACGAAATGATTAATGATGATGAAATGAACTG 10320
 Db 2328 AspIleGlnAsnAspGlyIleProSerSerLysIleThrAspAsnGluTyrAsnGlnLeu 2347
 QY 10321 AAACATGATTTGATGCTCAATTTA ---CCAAATACAGACCAAAATATATAT 10371
 Db 2348 LysAspLeuPheIleSerGlnTyrLeuGlnSerGluProAsnThrGluPro ---Asn 2365
 QY 10372 TACAAAGTCCAGTATTCACATGATACAGAACCTTAATCTTATTTGATTAATCCT 10431
 Db 2366 MetLeuGlyTyrAsnValAspAsnAsnThrHisProThrThrSerHis ---HisAspVal 2384
 QY 10432 GAAGAAAAACCTTTATATATGATGATGATGATGATGATGATGATGATGATGATGAT 10491
 Db 2385 GluGluLysProPheIleMetSerIleHisAspArgAsnLeuPheSerGlyGluGluTyr 2404
 QY 10492 AGTTAT ---AATTTAATATGATGATGAT 10518
 Db 2405 AsnTyrAspMetPheAsnSerGlyAsnAsnProIleAsnIleSerAspSerHisSer 2424
 QY 10519 ---ACTAATATGATATTCGAATGAATGCTGGAATGATTTGATATAGA 10563
 Db 2425 MetAspSerLeuThrSerAsnAsnHisSerProTyrAsnAspLysAsnAspLeuTyrSer 2444
 QY 10564 GGTATGATTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10623
 Db 2445 GlyIleAspLeuIleAsnAspAlaLeu ---SerGlyAsnHisIleAspIleTyrAspLys 2464
 QY 10624 TATTTG 10628

Db 1204 AsnProProProGlnThrLysAspAlaCysIleCysAsnThrArgProProLysGlu 1223
QY 3736 GTACAGATTAAGCAAGAAAAA-----AAAGCG 3762
Db 1224 AsperArgLysArgSerGluAspSerAspGluGluLysValLysVal 1243
QY 3763 GAAGAAAAAGATACGCA-----3780
Db 1244 GluGluLysAlaThrGluAspAlaValAspThrGlyProProProAlaProLysGluAla 1263
QY 3781 -----TGTAAACAGCATGATGATTAAGTAAAGCAAGCAATGCAAG 3822
Db 1264 ThrThrThrLeuAspValCysProIleValAlaGlyValLeuThr-----Lys 1279
QY 3823 AAACAACTAGACAGATTGTCATCCAAAAAAGCAATGATGATATCCC---GATTGGCA 3879
Db 1280 GluAsnLeuGluAsnAlaCysProThrLysTyrGlyProLysAlaProThrSerThrLys 1299
QY 3880 TGC-----GGAATATATA 3891
Db 1300 CysIleProThrGluLysThrAsnAlaAlaThrGlySerGluGlySerSerGlyAsnGly 1319
QY 3892 AATTAGTGAACACCCCTGCT-----3912
Db 1320 AlaLeuGluArgAlaLysArgAlaThrValGluSerGlySerProValThrSerAsnSer 1339
QY 3913 -----GTGTGATGCCCTTAGACACAAAGTATGCGTA-----CATTTCTTG 3957
Db 1340 GlySerIleCysIleProProArgArgArgArgLysIleGluLysLeuHisAspThr 1359
QY 3958 GCAAATGATATGAATATAAATAAATTAACATCAAGTT-----3996
Db 1360 AlaSerGlyAsnThrValValSerGlyGluAlaGlnThrProGluGlyLysThrSerSer 1379
QY 3997 -----AATTTAAAGAGCTTTCATCAAAATCTGACCA 4029
Db 1380 ProSerGlyLysGluThrProSerAspLysLeuArgThrAlaPheIleGlnSerIleAla 1399
QY 4030 GCAGAAACATTCCTTCATGATATTAATAAAGTAAAGATGATGAGAAAGAAATCAATC 4089
Db 1400 IleGluThrPhePheLeuThrAspArgLysLysGlu-----LysGluIle 1415
QY 4090 GATTAAGATTAATAAAGAA-----4107
Db 1416 GluLysLysGluLysValAlaAlaAsnGlyLysLeuValProSerLeuAsnGlyPro 1435
QY 4108 -----GGCAAA 4113
Db 1436 ProGlnGlnProGlyValThrGlyAspSerProGlnSerLysLeuGlnGlnThrGlyVal 1455
QY 4114 ATTCCTCCGCAATTTTGGATTCATTCATACATTTGGAGATTTAGATTTTGA 4173
Db 1456 IleProProPheLeuArgGluMetPheThrLeuGlyAspTyrAlaAspIlePhe 1475
QY 4174 TTGGACACAGT-----ATATCAAAAGCATAGGTGAG-----4206
Db 1476 PheGlyLysAsnAspIleValIleAspThrLysAsnGlyLysAspIleAlaGluArg 1495
QY 4207 GGAAGTAAACATAAAGCAATAGATTCCTTTCAAAATAGTGACCAAAATCTCCT 4266
Db 1496 GluLysLysIleLysAspAlaIleGluArgValLeuLysAsnAlaAspSerGlnProPro 1515
QY 4267 AATGAAAAACACGCAAGATGTCGACAGACATAGTCATGATGATGAGAAAGTATG 4326
Db 1516 SerAspLysArgGlnThrThrProGluGlnAsnGlyLysIleThrPheAsnGlyMet 1535
QY 4327 CTATGTCCTACTAGT-----AAATTT 4347
Db 1536 IleCysAlaLeuThrThrLysGluLysAspGluLysGlyThrProLeuLysGlnAsnGlu 1555
QY 4348 GGGGCAAAA-----AAAGATGATTTTACCGAAACCTACGCT 4383
Db 1556 GlyLeuLysSerAlaLeuThrAspGluLysAsnLysLysProLysAspGlnLysTyrGln 1575

QY 4384 TACAAACAGTCAAAATTTAGTCAAAAGC-----4413
Db 1576 TyrAspLysValLysLeuAspGluAsnSerGlyThrSerProLysThrAsnAspHisVal 1595
QY 4414 -----ACCACCTTGGAGAAATTTGCCAAACAGACCCAGTTTTCAGATGGCTAACCGAA 4467
Db 1596 ProProThrProLeuThrAsnPheIleSerArgProProTyrPheArgTyrLeuGluLys 1615
QY 4468 TGTACAGCAGTATGCTATACAGACAAAAAATTTGAAGAGATGACGAGAAAAATGT 4527
Db 1616 TrpGlyGluThrPheCysArgGluArgLysLysArgLysValGluLys 1635
QY 4528 AAGTCAATGAC-----CAATTGAATGTGAT-----4554
Db 1636 MetAspGluAspGlyLysLysGlnLysCysSerGlyAspGlyLysAspCysGluIle 1655
QY 4555 -----ACGAATGTAATTAAGAAATGC 4575
Db 1656 ArgLysGlnAspTyrSerThrValArgAspPheTyrCysProGluCysGlyLysTyrCys 1675
QY 4576 GAGCCTACGTTAAATATATGAAAAA-----GAGTGCATTCACAGATTAATAT 4632
Db 1676 ArgPheTyrLysArgTyrIleGluLysLysAspGluTyrAspLysGlnLysGluAla 1695
QY 4633 TACAAGATGAAAGCGACAAAAAAGATTCATGACACACATGCTGTAATGTTACA 4692
Db 1696 TyrAsnAsnGlnLysThrAspAlaArgArgAsnAsnAspAsnAlaPheSerThrThr 1715
QY 4693 GACTATACGTGAAAGCAATGCAACAGATTAATCTGTAACAGAAATTTACTGTAATGCT 4749
Db 1716 LeuAspThrCysThrThrAlaGlyAspPheLeuGlnThrLeuLysAsnGlyProCysLys 1735
QY 4750 -----GGTGAATACCTGGAAGTCCCTGCTGCTGATACAAAGAAATATACAAATG 4797
Db 1736 AsnAspAsnValAspAspSerGlyLysAsnLysLysIlePheAspGluAsnGlyAspThr 1755
QY 4798 TTGAAGAAACAGGCTTCTATGATGCGCAGCAAAACATGTGGGGGACCAAAATTTATGAA 4857
Db 1756 PheLysTyrThrGlnTyrCysGlyThr-----CysSerLeuAsnGlyPheLysCys 1772
QY 4858 AATGACACAAA-----TATACTACATTTGCAATGAATGAATGCAAAAGATTA 4908
Db 1773 AsnGlyAspAspCysArgValArgThrAsnValThr-----CysAsnLysSer 1788
QY 4909 GTAAGAGAGCAAAACACAGGTGCTATTAAAGTGCACAAACAAAGCTCTAATACCTCAAT 4968
Db 1789 AsnArgThrThrThrIleThrAlaAspAspIleLysAsnGlyLys-----AsnSerAla 1806
QY 4969 AACTGAAAGAAATGACAGATGATGCTTTTCTTCCTGCTGACATGATATGTTT 5028
Db 1807 GluIleAsnMetLeuValSerAsp-----1814
QY 5029 CATGCAATGATGGCAATATATACAGATCCAGAAAGTAAAGATGAAGATGGTTCGCAAA 5088
Db 1815 -----AspIleAsnSerGlyAsnGlyPheAsn---1823
QY 5089 AGATTGATGAAAGTGGCGCAACGAAGGATGCAATTTGGTCAATATCAACAAAGAAAA 5148
Db 1824 -----AspLeuGluLysLysAsnAlaAsnIlePheLysGlyLysGluLys 1840
QY 5149 AAAGAAAAAGAGAAATTAATAAAGTCGGATGCGCACAAATATCTTAAAGAGGTCCGCT 5208
Db 1841 LysTyrLys-----CysValTyrPheCysLysSerSerAspAla 1852
QY 5209 TGTAGTCAATGAAATATATGTTTATGATTAAGATATATATCTA-----5256
Db 1853 CysGlyLeuLysLysAsnAspIleAspGlnAsnGlnIleIleLeuIleArgAlaLeu 1872
QY 5257 -----GTAATGATTAATTTGGAAGATGAAACAAAGACCGAGAAATTTGAG 5307
Db 1873 PheLysArgTyrPheLysGluTyrPheLeuAspAspTyrAsnLysIleArgLysLysLeuAsn 1892

```
QY 5308 AAAATATTACAAAAATGACATGCTGGCAAGAGTATAGTACTACAGCAAT 5367
Db 1893 ProCysIleAsnAsn-----GlyGluLysAlaIleCysThrAsnGlyCys 1907
QY 5368 CCCGGTAGTACTCGCGAAATTTTCTGGACGAAATATAGCAATGTGTGNAAGCA 5427
Db 1908 ValGluGln-----TrrIleAsnGlnLysArgThrGluTrpThr--- 1920
QY 5428 ATGATATGGGGGTACAAACGTGTAGGATGATGAAATATAGTGAAGAAGT 5487
Db 1921 -----AsnIleLysAsn-----ArgPhe 1926
QY 5488 GATGAAGATCTAAAAAATGTGTTCTGTACCTTCAGATGATGATTCCTATGGGAAA 5547
Db 1927 AsnGluGlnTyr-----AsnGlyAspAspThrGluMet---Lys 1938
QY 5548 AATCGGATGAAGTACTGCGTATCATGTTCTTCGATGGTTTCCGAAATGGGGTGAAGAT 5607
Db 1939 SerSerPheArgSerPheLeuValAspLeuIleArgGlnIleAla-----AlaThr 1955
QY 5608 TTTTGCAACATTAAGAAAAAGAAATGAGAAATTTGTAGGGCG-----TGTAAATGAT 5661
Db 1956 IleAspLysGlyAsnHisAsnGlyLeuValLysLeuValLysSerValLysCysAsn--- 1974
QY 5662 TATACTTGTGGTATATGAGATTAAGAAAGAAATGTACAGATGCGGTACACATAT 5721
Db 1975 -----CysGlyAsnAsnSerGlnAsnGlyLysGlu----- 1984
QY 5722 AAAAAATTATGTAGTGAAGTGAACACAGTATGAAACAAATCAAAAAATATGGTGA 5781
Db 1985 -----GlyGlu 1986
QY 5782 AATTAAGACAAATATATTCGAGATCCTGTGGCAAAAGATGACAGAGCGCTCGGAA 5841
Db 1987 GluAsnAspLeuValLeuLys-----LeuLeuGln 1996
QY 5842 TATTAGACAAACATTAATAAAATTTGTGAAT-----AAAAGTCAGATTTGTGAA 5895
Db 1997 LysLeuGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 2016
QY 5896 TATTAAGTATGTAAGATGTGTCCACACAGCGATTAAGTAT----- 5937
Db 2017 GlnProCys-----GluValSerProAsnHisIleGluAspLeuGlnProLeuGlu 2034
QY 5938 -----GTAATAGTCAAAATATATGCCGATCATTAAGCAT-----GAACCA 5979
Db 2035 GluGluGluAsnThrValGluHisProLysIleCysAspAspValLeuLysThrGlnPro 2054
QY 5980 AAAGAAAGTGAAGAAAGTGAATTTGTCAAGTCCACAGGTCCACACGCTGTACGAAAG 6039
Db 2055 GlnProGluGluProGluLysLysLysLysLysLysLysLysLysLysLysLysLys 2074
QY 6040 GAATACACCGTACACGCGGTATCATATATCAAAAGCGACGCGCATGCAAAAAGAAAGCG 6099
Db 2075 GluGluGluGluGluLys-----LysGluGluLysAspLysGluGluGlu 2089
QY 6100 AAAACAGCGCGCGCTCAAAACAGCGGAAAAAGTGAATCTTAACACAGAAATGCGA 6159
Db 2090 ProGlyLeuProProProProProProProProProProProProProProProPro 2108
QY 6160 GCACAAACAGAAACCGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 6208
Db 2108 opProLysProLysProProProGlnValGluLysAsnProTrpGluHis 2124
```

Search completed: April 28, 2003, 11:59:58
Job time : 1665.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 10:53:21 ; Search time 129 Seconds

(without alignments)
6834.270 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455
Sequence: 1 atgggggtctctctgcaata.....atatatgatgaagtattg 10628

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame-n2p.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US10087013/runat.28042003.102845.20254/app.query.fasta_1.10823
-DB=SwissProt.40 -OPMT=fastan -SUPER=n2p.rsp -MINMATCH=0.1 -LOOPEC=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -STRAT=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10087013.eccn_1.1.224.0runat.28042003.102845.20254 -NCPU=3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-NARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEX=7

Database : SwissProt.40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	618	3.2	1435	1	EBAL_PLAFC
2	435	2.2	2869	1	RBP1_PLAVB
3	420.5	2.2	6669	1	NEBU_HUMAN
4	384.5	2.0	3135	1	S230_PLAFO
5	379	1.9	3210	1	CENF_HUMAN
6	349	1.8	1153	1	PVDR_PLAKN
7	344.5	1.8	1070	1	PVDR_PLAKN
8	343	1.8	1875	1	MLPI_YEAST
9	337.5	1.7	1070	1	PVDR_PLAKN
10	337.5	1.7	1073	1	PVDR_PLAKN
11	329	1.7	2116	1	MYR2_DICDI
12	322.5	1.7	2748	1	NUM1_YEAST
13	319.5	1.6	1391	1	MSR2_DROH
14	313	1.6	2663	1	CENF_HUMAN
15	311	1.6	2492	1	ATRX_HUMAN
16	311	1.6	2704	1	ATRX_HUMAN
17	302	1.6	2476	1	ATRX_MOUSE
18	299.5	1.5	2339	1	RPCI_PLAFA

19	298.5	1.5	1956	1	ATX1_PLAFA	004956	Plasmodium
20	290	1.5	1957	1	YD86_SCHPO	010411	Schistosoma
21	289	1.5	1630	1	MSPI_PLAFC	P04932	Plasmodium
22	289	1.5	1639	1	MSPI_PLAFC	P04933	Plasmodium
23	289	1.5	1639	1	MSPI_PLAFC	P04934	Plasmodium
24	287	1.5	3418	1	BRD2_HUMAN	097592	Canis fam1
25	287	1.5	3685	1	DMD_HUMAN	P15587	homo sapien
26	279	1.4	3911	1	AKA9_HUMAN	P15582	homo sapien
27	274	1.4	2230	1	GOG4_HUMAN	099936	h a-kinase
28	271.5	1.4	1790	1	USO1_YEAST	013439	homo sapien
29	270.5	1.4	1805	1	HMW2_MYCE	P25386	saccharomyc
30	270.5	1.4	3678	1	DMD_MOUSE	P47460	mycoplasma
31	269.5	1.4	2022	1	ANT1_ONCVO	P15531	mus musculus
32	268	1.4	1658	1	YME7_YEAST	P21249	onchocerca
33	267.5	1.4	5430	1	ACF7_HUMAN	003661	saccharomyc
34	265	1.4	1679	1	YIO9_YEAST	09upn3	homo sapien
35	263.5	1.4	1726	1	MSPI_PLAFC	P40457	saccharomyc
36	262.5	1.3	5337	1	ACF7_MOUSE	P04934	Plasmodium
37	261.5	1.3	1726	1	MSPI_PLAFC	09qzr0	mus musculus
38	259.5	1.3	3660	1	DMD_CHICK	P15533	gallus gall
39	259	1.3	1557	1	DVAL_DICVI	024702	dictyocaulu
40	259	1.3	1727	1	ALM1_SCHPO	09utk5	Schistosoma
41	258	1.3	2871	1	DESP_HUMAN	P15974	homo sapien
42	255	1.3	1939	1	MYR6_MESAU	P13559	mesocricetu
43	252.5	1.3	1928	1	MYR1_YEAST	P08964	saccharomyc
44	245	1.3	2198	1	YIJ2_CAEEL	P34367	caenorhabdi
45	244.5	1.3	1701	1	MSPI_PLAFC	P13819	Plasmodium

ALIGNMENTS

RESULT 1
ID EBAL_PLAFC STANDARD: PRT: 1435 AA.
AC P19214:
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Erythrocyte-binding antigen EBA-175.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5835;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90377299; PubMed=2204835;
RA Sim B.K.L.;
RT "sequence conservation of a functional domain of erythrocyte binding
RT antigen 175 in Plasmodium falciparum".
RL Mol. Biochem. Parasitol. 41:293-296(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: X52524; CNA36756.1; -
KW Antigen.
KW DOMAIN 159 1104
FT
FT DOMAIN 159 1104
FT
FT VARIANT 1031 1031 ERYTHROCYTES.
SQ SEQUENCE 1435 AA; 167389 MW; 32A4309021B1C3D6 CRC64;
E -> V (IN STRAINS FCR-3 AND ITG).
Alignment Scores:
Pred. No.: 5.4e-22
Score: 618.00
Percent Similarity: 36.80%
Best Local Similarity: 22.20%
Query Match: 3.18%
Length: 1435
Matches: 339
Conservative: 223
Mismatch: 528
Indels: 42
Gaps: 72

US-10-087-013-1 (1-10628) x EBA1_PLAFC (1-1435)

OY 3301 GAAGCTTGAATGATATATATAGATATATGGAAGAACAA-----TCG 3351
 Db 20 LysAlaArgAsnGluTyr--AspIle-----LysGluAsnGluLysPheLeu 34
 OY 3352 AATATATATCATGATTAATACAAAGATTTACATGACAAAGCAAAATGCTGTAGTAAT 3411
 Db 35 AspValTyrLysGluLysPheAsnGluLeuAspLysLysTyrGluValGluLys 54
 OY 3412 AGTGGTATGAGCTTCCAGTACTGCCAAAATCATATGACAGGAATGTTATGAATT 3471
 Db 55 Thr-----AspLysLysIlePheThrPhe 62
 OY 3472 TTG--TCGGAATTTATACCAAAATGCTGGCAAAATGATTAAGT----- 3516
 Db 63 IleGluAsnLysLeuAspIleLeuAsnAsnSerLysPheAsnLysArgTyrLysSerTyr 82
 OY 3517 GGTACTAGTGTGAAAAGCGCTGATGGTACTGACTACACACCGATGAAATGTTGGAGCA 3576
 Db 83 GlyThrProAspAsn-----IleAspLysAsnMetSerLeuIleAsn----- 96
 OY 3577 TATCTCATGATACGAGAAATTTGATGATTTGTCAGTCACAAATGAGTTTGATGAA 3636
 Db 97 ---LysHisAsn--AsnGlu---GluMetPheAsnAsnAsnTyrGlnSerPheLeuSerTh 114
 OY 3637 AAAAGTATGATGTAAGCATTAACGAAAAATATGCTTTAGACATTAACCCACAGCAATGAT 3696
 Db 114 rSerSerLeuIleLysGlnAsnLysTyrValProIle-----AsnAlaValArgVal 131
 OY 3697 GGTGCGTGTGTTGTAATAATGATGCAACGCAAGGTTACAGATTAACGAAAAA 3756
 Db 131 LSerArgIleLeuSerPheLeuAsp----- 139
 OY 3757 AAAGCGAAGAAAAGCATACGGAATGTAACAGTG-----ATGATATACTTAA 3806
 Db 140 -----SerArgIleAsnAsnGluLysArgAsnThrSerSerAsnAsnGluValLeu 156
 OY 3807 AGAA---AACGATGGAAGAAAGAACAGTAGAA---GATTCATCCAAAAAGCATGTAA 3860
 Db 156 rAsnCysArgGluLysArgLysGluLysMetLysTrpAspCysLysLysLysAsnAspArgSe 176
 OY 3861 TGGATATCCCGATTTGGCAATGCGGAATATTAATTTAGTGAAGACCCGCTGTGTAT 3920
 Db 176 rAsnTyr-----ValCysIle 181
 OY 3921 GCCCCTAGAGACAAAGTATGCTATGCTATTTGCGCAAAATGATTAAGAAATAAAA 3980
 Db 181 eProAspArgArgIleGlnLeuLysIleValAsnLeuSer-----IleIleLys 197
 OY 3981 ATTACATCACAAGTTAATTTAAAGACGTTTCATCAATCTGACAGCAGCAAGAACAT 4040
 Db 197 sThrTyrThrLysGluThrMetLysAspHisPheIleGluAlaSerLysLysGlu----- 215
 OY 4041 CTTTCATGCTGTTTATTTAATAAGTAGATGTTGAAGGAATGAATCGATTAAGAATT 4100
 Db 216 ---SerGlnLeuLeuLysLysAsnAspAsnLysTyrAsnSer----- 229
 OY 4101 AAAAGAGGCAAAATTCCTCCCGCATTTTGGATGCCATGTTCTACACATTTGGAGATTA 4160
 Db 230 -----LysPheCysAsnAspLeuLysAsnSerPheLeuAspTyr 242
 OY 4161 TAGAGATTTTATTTGGAACAGATATATCAAAAGTCATGCTGAGGAGCAAGAACTAAA 4220
 Db 242 rGluHisLeuAlaMetGluAsnAspMetAspPheLys---GlyTyrSerThrLysAlaGlu 261
 OY 4221 AGAGCAATATGATTTCTCTTTCAAAATAGTGACCAAAATCTCTTAATGGAAGAAACA-- 4278
 Db 261 uAsnLysIleGlnGluValPheLysGluAlaHisGluLysLeuSerGluHisLysIleLys 281
 OY 4279 -----CGCCAGAGATGTGAGACAGACATAGCATGAGATATGGAAGCTATGCTATG 4331
 Db 4331 ----- 4331

Db 281 sAsnPheArgLysGluThrTrpAsnGluPheArgGluLysLeuThrGluAlaMetLeu-- 300
 OY 4332 TGCAGTAGTAAAAATGGGCAAAAAAGATGATTTTACGGAATACAGGTACACAA 4391
 Db 301 -----SerGlnHisLysAsnAs 306
 OY 4392 CGTCAAATTTAGTGACAAAAAGCACCCTTTGGAGAAATTTGCCAAAGCAAGCTTTT 4451
 Db 306 nIleAsn--AsnCysLysAsnIleProGlnGluGluLeu-----GlnIleTh 321
 OY 4452 ACGATGGCTAACCGAATGAGTACAGCAGCTATTGCTATACACGCAAAATATTATTAAGA 4511
 Db 321 rGluThrIleLysGluThrPheIleGluThrPheLeuGluGluArgAsnArgSerLysLe 341
 OY 4512 TGTGCGAGAAAAATGTAAGTCAAAATGACAAATGGAAG---TGTGATACAGAAATGA 4568
 Db 341 uProLysSerLysCysLysAsnAsnThrLeuTyrGluAlaCysGluLysGluCysIleAs 361
 OY 4569 GAAATGCGAGGACTACGTTAATAT--ATGAAAAAAGAGTGAGATTCCACAAAGA 4625
 Db 361 ProCysMetLysTyrArgAspThrIleLeuArgSerLysPheGluThrPheIleuSe 381
 OY 4626 TAAATATTACAAGATGAAAGCGACAAAAAAGATTGATAGACAAACATTTGCTGAT 4685
 Db 381 rLysGluTyrGluThrGlnGluValProLysGlu----- 392
 OY 4686 GGTACAGACTTACTGGAACAGATGCAACAGATTACTTGAACGAAATTTACTGTAG 4745
 Db 393 -----AsnAlaGluAsnTyrLeu---IleLysIleSerGluAs 404
 OY 4746 TTGTGCTGATAGCTGGAGAGTCCCTGTGTGACAAAGAAATATCAATTTTATGA 4805
 Db 404 nLysAsnAsp-----AlaLysValSerLeuLeuAs 415
 OY 4806 ACAGGCTTACTATGATGCCGCAAAACATTTGGCTGC-----ACAAAAATTATGA 4856
 Db 415 nAsnCysAspAlaGluTyrSerLysTyrCysAspCysLysHisThrThrIleValLys 435
 OY 4857 A-----AATGAC-----GACAAATATCTAATCAATTTGACAGTAA 4889
 Db 435 sSerValLeuAsnGlnLysAsnAspAsnThrIleLysGluLysArgGluHisIleAspLeuAs 455
 OY 4890 AGATAGTGCAAA---GGATAGTAAAGAGGCA---AACACAGCTCTATTAGTGCA 4943
 Db 455 rAspPheSerLysPheGluCysAspLysAsnSerValAspThrAsnThrLysValTrpGlu 475
 OY 4944 AAACAAAGCTCTATATACATCAATCAATCAATGCAAGATTCAGCAAGCTCTTTCC 5003
 Db 475 uCysLysAsnProTyrIleLeuSer-----ThrLysAspValCysValPr 490
 OY 5004 TTCTGCTGAGCTACGATATATGTTTTCATGATTCGATTCGATTCGATTCGATTCGAT 5063
 Db 490 oProArgArgGlnGluLysLeuLysGlnLysGlnLysAsnIleAspArgIleTyr----- 505
 OY 5064 TAAAGATGAAATGGGTTG-----GGAAGAAGATTTGAAGATGGCGCAAGG 5117
 Db 506 ---AspLysAsnLeuLeuMetIleLysGlnHisIleLeuAlaIleValIleTyrGluSe 524
 OY 5118 GTACAAATTTGGGCAATACATACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCTGCA 5177
 Db 524 rArgIleLeuLysArgLysTyrLysAsnLysAspAspLysGlu----- 538
 OY 5178 TGGCGCAAAATATTCTTAGAGGTCCCGCTGTAGTGTATGAATATATGTTTATGA 5237
 Db 539 -----ValCysLysIleIleAsnLysThrPheAlaAs 549
 OY 5238 TTTAAGATATATCTAGTATGATTAATTTGGAAGATGAAAGAAAGCAAGCAGAGA 5297
 Db 549 rIleArgAspIleIleGluCysLysThrAspTyrTrpAsnAsp-----LeuSerAsnArg 566
 OY 5298 AAATTTGAAGAAATATTTAACAAGAAATGCAATGATGTCGAAAGAAAGAAAGATAGTAC 5357
 Db 566 gLysLeuValGluLysIleAsnThrAsnSerLysTyrValHisArgAsnLysLysAsn-- 585

```
QY 5358 TACAGGAATCCGGTAGTACTCGCGAATAATTTCTTGGACGAAATAAGAAATGT 5417
      :::::
Db 586 -----AspLysLeuPheArgAspGluTrpTrpLysValIleLysLysAspVa 601
QY 5418 GTGAAGCAATATATCGCGGTACAAACGTGGTAGGATGATGAAATAGTGAATAG 5477
      |||||
Db 601 LTrpAsnValIleSerTrpValPheLys----- 610
QY 5478 TGCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5537
      |||
Db 611 -----AspLysTh 613
QY 5538 TATGGGAAAAATCGCGATGAGTACTCGTATCAGTTTCTTGATGTTGCCGATG 5597
      :::::
Db 613 rValCysLysGluAspAspIleGluAsnIleProGlnPheArgTrpPheSerGluTr 633
QY 5598 GGGTGAAGATTTTGCACATTAAGAAAAGAAATGGATGGATGGAGGGCGGTAA 5657
      |||||
Db 633 pGlyAspAspTrpCysGlnAspLysTrpLysMetIleGluThrLeuLysValGluCys 653
QY 5658 TGATTTACTTGTGTATGATGATGATGATGATGATGATGATGATGATGATGATG 5717
      :::::
Db 653 sGluLysProCysGluAspAsp-----AsnCysLysSerLysCysAsn 668
QY 5718 ATTTAAAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATG 5777
      |||||
Db 668 rTyLysGluTrpIleSerLysLysGluGluTrpAsnLysGlnIleLysGlnTy 688
QY 5778 TGGG-----AATTAAGCAAAATATATTCGACATCCCTGTGCAAAAGATGC 5825
      |||
Db 688 nGluTyLysGluLysAsnTrpLysMetTyLysSerGluPheLysSerIleLysPro 708
QY 5826 AGAGAGCGCTCGGATATTTAGACAACAATTAATAAATTTGCAAAATTAAGTGG 5885
      :::::
Db 708 uValTyLysLeuLysLysSerGluLys-----CysSerAsnLeuAsnPh 723
QY 5886 AGATTGTAATTAAG-----TGT-- 5904
      :::::
Db 723 eGluAspGluPheLysGluGluLeuHisSerAspTyLysAsnLysCysThrMetCys 743
QY 5905 -----ATGAAAGATGTGTCACACAGCATTAAGTATGATGATGATGATGATG 5960
      :::::
Db 743 oGluValLysAspValProIleSer-----IleIleArgAsnAsnGluIleThrSerGln 762
QY 5961 ATCATTTAGACGATGACCAAAAGAAAT-----GAAGCAAGTGTATTTGCA 6008
      :::::
Db 762 uAlaValProGluGluAsnThrGluIleAlaHisArgThrGluThrProSerIleSer 782
QY 6009 AGTGCACGAGGTCCACACGTGTAGAGGAAACACCGTCACACGCGGTATCATGT 6068
      :::::
Db 782 uGlyProLysGluLysnGluGlnLysGluArgAspAspSer-----Le 797
QY 6069 ATCAAAAAGCAGCATGCAAAAAGAAAGCAAGCGCGCTCAAAAAGCGCGGA 6128
      :::::
Db 797 uSerLysIleSerValSer-----ProGluAsnSerHisPro-- 809
QY 6129 AAAAGTGAATCTTACAAACAGAAATGCGACACAACAGAACCCGACAGCAGCA 6188
      :::::
Db 810 -----G 810
QY 6189 ACAACACGAAAGCAATTCACACAGCAACACAGAAATTCGACTGGC---ACAAT 6245
      :::::
Db 810 uThrAspAlaLysAspThrSerAsnLeuLysLysLysValAspValAspIleSer 830
QY 6246 GGTAAAGCATCTTCTTCGATTAACACAGATAGCAGGGGTGAATAGAGGTGTATCC 6305
      |||||
Db 830 LProLysAlaValIleLysSerProAsn-----AspAsnIleAsn-- 844
QY 6306 AAAAAGCTATGACAAATATCTTAATGGGGTGTATTTAGTGAAGTCTAAAGAAATGA 6365
      :::::
Db 845 -----ValThrGluGlnGluLysAspAsnIle 852
QY 6366 AATGCGATATGATGCTCTTACGAGCAAAAATTTATGATTAATATATATATATTT 6425
      :::::
Db 852 eSerGluVal-----AsnSerLysProLeuSerAspAspValArgProAs 867
QY 6426 AATTTATGACACGTGAATAAGCGGTGACATGATGATTAAGAGCGCTTTATTAATGTC 6485
      |||
Db 867 pLysLysGluLeuGluAspGlnAsnSerAspGluSerGluGluThrValVal----- 884
QY 6486 AGCAATAGAACTCAATTTTGTGTTAAATATATATATGAAATTCGACAGCAAAA 6545
      :::::
Db 885 -----AsnHisIleSerLysSerProSerIleAsnAs 895
QY 6546 TGAATTCGAAATGGAACAATTCACAGATGAATTTAAAGATATGATTAATACATATG 6605
      :::::
Db 895 nGlyAspAspSerCys-----GlySerAlaThrValSe 907
QY 6606 TGATTTAAGATATGTTTTTGGAACTGATATTTCTATGATTAATAAATATATATCT 6665
      :::::
Db 907 rGluSerSerSerSerAsnThrGlyLeuSerIleAspAspAspArgAsnGluAspThrPh 927
QY 6666 AACAAATAGTGTAAACAACATTCATCAATGAATTAATGAATAAACAAGATGAATAA 6725
      :::::
Db 927 eValArgThrGlnAspThrAlaAsnThrGluAspValIleArgLysGluAsnAlaAsp 947
QY 6726 AGATGAAGAAATTCAGTAAATATTTTGGAGCAAAAATTAATTTTATTTGGCAGGAT 6785
      |||||
Db 947 sAspGluAspGluLysGluLysAlaAspGluLys----- 958
QY 6786 GATATATGATTAATCATATCATCTCACAGACAGAAAC-----GAAAAGAAA 6833
      |||
Db 959 -----HisSerThrSerGluSerLysSerProGluGluLys 972
QY 6834 AATTAAGATTAATTAACAGTACATGATGATGATGATGATGATGATGATGATGAT 6893
      :::::
Db 972 LLeuThrAspAsnGluLysGluLysAsnSerLeuAsn-----HisGlu 986
QY 6894 TGTAAAAAGCCCCAATTTTGGATGTTGCACGAATGGCGACAGCAATTTTGTAA 6953
      |||||
Db 986 uValLysGluHisThrSerAsnSerAspAsnValGlnGlnSerGluLysIleValAsn 1006
QY 6954 GAGCAAGCAACAGTTGTAA-----TTGAGCGCGG 6986
      :::::
Db 1006 LAsnValGluLysGluLeuLysAspThrLeuGluAsnProSerSerLeuAspGlu 1026
QY 6987 CTGTAAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7017
      |||
Db 1026 yLysAlaHisGluGluLeuSerGluProAsnLeuSerSerAspGlnAspMetSerAsnTh 1046
QY 7018 -----GACGTAAGACACAAAGATGTCAGAGCGGTGTAAACATATCAAA 7064
      :::::
Db 1046 rProGlyProLeuAspAsnThrSerGluGluThrThrGluArg-----IleSerAsnAs 1064
QY 7065 TTTATTAAGAGTGAAGAACTGAATGAAGCAAAAGCAAAAGTTCAAAAGAGAT-- 7122
      |||
Db 1064 nGluTyLysValAsnGluArgGluAspGluArgThrLeuThrLysGluLysAspIle 1084
QY 7123 -----AAAGATGCGCAAAAAGTATAGATTTATCC 7151
      :::::
Db 1084 eValLeuLysSerHisMetAsnArgLysArgLysAspGluLysLeuTyArgGlu----- 1102
QY 7152 TTTACTGAAAGACATAGGAAGGCAACATGTGCTCATGATATTTAAACATGAAT 7211
      :::::
Db 1103 -----AsnSerAspLeuSerThrValAsnAspGluSerGluAspAlaGluAlaLysMe 1120
QY 7212 AAAGAAATTAATGTCGATTAAGATTTCTTGTATGCAAAAACCTTCTTCAACAATAC 7271
      |||||
Db 1120 tLys-----GlyAsn--AspThrSerGluMetSerHisAsnSerGlnHisIle 1136
QY 7272 AAAAACAACAACAATCACAATCATCGATGCTAATGATATGCCAGAAATCGCTGAT-- 7329
      :::::
Db 1136 eGluSerAspGlnGlnLys-----AsnAspMetLysThrValGlyAspLe 1151
QY 7330 -----TATGTCTCTGAAGATTTAACAGTGTGAGTGTCTGTAACCTTCAAAA-- 7377
```

Db 1151 uGlyThrThrHisValGlnaGlnIleSerValProValThrcylGluIleAspGlu 1171
 QY 7378 -----AAGGATCTATGATCATACAAAAAATTAAGTAAATAC 7424
 Db 1171 sleuArgIleSerLysGluSerLysIleHis-----LysAlaGluGluGluAspGlu 1189
 QY 7425 TATGATATGCTAGAGAAAGCAATATTAATCTAAAGACAGAAATAATATGGA 7484
 Db 1189 rHisThrAspIleHisLys-----IleAsnProGluLysPargAsnSerAsnTh 1205
 QY 7485 TATTACCTTGAAGAAAAATTTATACCTATGATCTACAAAGAAAGAAAGTAAAA 7544
 Db 1205 rLeuHisLeuLysAsp-----IleArgAsnGluGluAsnGlu 1218
 QY 7545 TAGTTGACATATATATAT 7563
 Db 1218 gHisLeuThrAsnGlnAsn 1224

RESULT 2

PL PLAVB STANDARD; PRT; 2869 AA.
 Q00758;

DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Reticulocyte binding protein 1 precursor.
 GN RBP1.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 mercoletes";
 RL Cell 69:1213-1226(1992).
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 HUMAN RETICULOCYTE CELLS.
 CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

DR EMBL; M88097; AAA29743.1;
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17
 FT CHAIN 18 2869
 FT DOMAIN 18 2807
 FT TRANSMEM 2808 2826
 FT DOMAIN 2827 2869
 FT SITE 1030 1032
 FT SITE 2599 2601
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;
 FT SIGNAL 1 17
 FT CHAIN 18 2869
 FT DOMAIN 18 2807
 FT TRANSMEM 2808 2826
 FT DOMAIN 2827 2869
 FT SITE 1030 1032
 FT SITE 2599 2601
 SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Alignment Scores:

Pred. No.: 1,78e-13 Length: 2869
 Score: 435.00 Matches: 529
 Percent Similarity: 33.38% Conservative: 454
 Best Local Similarity: 17.96% Mismatches: 978
 Query Match: 2.24% Indels: 984
 DB: 1 Gaps: 129

US-10-087-013-1 (1-10628) x RBP1_PLAVB (1-2869)

QY 40 AATGACATCATCATAGAGGAGATGTAAAGCCCTATATATAAGAAAGTCACAA 99

Db 428 AsplAlaLalLysLysLeuGluSer-----IleAsnGluLysGlu 441
 QY 100 AGTCAAGAAATGTTTGGAACTTATGCCAAATATAGACAT---CCATCAAAATAT 156
 Db 442 SerAlaGluLysLeuSerGluLysPheGluAspAsnSerLysAlaLeuAlaAsnSerLys 461
 QY 157 GCA---AAGACATGTCGATTCGTTGAAAGGGATTGACGAAAGCAAAATTCGTGT 213
 Db 462 CysIlePheGluIleLysThrLeuAsnGluProIleLysLysAlaLys----- 478
 QY 214 GGTCTTACGCCGCTAATATAGATATATATATATATATATATATATATATATATAT 273
 Db 479 -----GluSerLysValIleLysSerAsnGluLeuLeuSerThrIleLeuAspThr 496
 QY 274 CATAGCAACATCTAATTTACGTATGATGATGATGATGATGATGATGATGATGATGAT 333
 Db 497 GlyLysSerAlaThrAlaLeu----- 503
 QY 334 AGAAGAAACAAACGCTTATGATGATGATGATGATGATGATGATGATGATGATGAT 393
 Db 504 -----GluGluSerThrPheAspGlu-----GluCys---AsnLysIleLysThr 518
 QY 394 -----TATTAAGAAAGAAATGATGCTATAGCCGTGGCCACCTAGAAAGCATATG 447
 Db 519 GluAlaGluLysValLysAspAspAla-----GluAspIle 530
 QY 448 TGTGATTAAGAAC-----TTGAAGCTCTAATGATTAATATATATATATATATAT 486
 Db 531 CysGluLysAsnGluGluIleLysThrLysLeuProGluSerGluAspGluThrIleAsp 550
 QY 487 CAATAATTCATGATTTATGGAATATGATGATGATGATGATGATGATGATGATGATGAT 546
 Db 551 AspLysIleAsnAspLeuGluLysLeuLysLeuLysLeuLysLeuLysLeuLysLeuLys 569
 QY 547 ATGTGATTAATTCATTCACATTAAGAAAGAACTTCAGACGCTGCTGCTGCTGCTGCT 606
 Db 570 IleValAsnAsnSerGluPhe-----IleSerAsnArg 580
 QY 607 TTTCGATATATAGGATATGATATGTAAGAGATATATATATATATATATATATATATAT 666
 Db 581 TyrLysAsnIleLysGlu-----AsnLeuLysGlu 590
 QY 667 AAGTAGAAACGGGTCTCCGAGAGTTTCAGAAATATACATGATGATGATGATGATGATGAT 726
 Db 591 ThrTyrGluThrGluLeuAsnAspIleGlyLys----- 601
 QY 727 GTAAAAATGATTTCAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 786
 Db 602 LeuGluLysAsp-----ThrSerLysValAsnPheTyrLeuMetGluIleArgLys 618
 QY 787 TGAATGTGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAATGAAT 843
 Db 619 IleAsnThrGluLysThrLysIleAspGluSerLeuGluThrValGluLysPheTyrLys 638
 QY 844 TCT----- 846
 Db 639 GluIleLeuAspSerLysGluLysIleTyrGluLeuLysIleGluPheGluLysSerVal 658
 QY 847 -----GGATATTTTATGCAATCAAGAAAGTAAATACCATATTTTCAATCTTAA 897
 Db 659 ThrGluIleAsnArgLeuGluAspGlyGluSerAlaArgAspLeu----- 673
 QY 898 TGGCGCCATTAACAA-----GGAAGGTTCCCTACCAATTTAGATAT 939
 Db 674 -----HisGluGluGluLysGluIleLeuAspLysMetAlaLysLysValHisLys 691
 QY 940 GTCCTCATATTTACCT-----TGGTTCGACGATGGGAGAGA 978
 Db 692 LeuLysGluLeuLeuSerLeuLysGlyLysSerValTyrPheThrCyluMetAsnGlu 711
 QY 979 -----GAGTTTGGCGAAGAAATATTAATTAATTAAGAAAGTCAAGAAC 1023

Db 712 LeuLeuAsnThrAlaSerTyrAspAsnMetGluGlyPheSerAlaLysLysGluLys--- 730
 QY 1024 TCCTGCGAATGACAAAGACGCTTATATTGATCATTAATGACATGATTGTACGACA 1083
 Db 731 ---AlaAspAsnAspLLeuAsnAlaLeuTyrAsnSerValTyrArgLysAspLLeuAsnAla 749
 QY 1084 ACTATTGG-----AAAAAGCTATTTGGCTTTG--- 1113
 Db 750 LeuIleGluGluValGluLysPheValThrGluAsnLysGluSerThrLeuGluMetLeu 769
 QY 1113 ----- 1113
 Db 770 LysAspGluGluMetGluLysLeuGluAspAlaLysGluThrPheAlaLysLeuAsn 789
 QY 1114 -----GATATAAGTACTGACTGCTGCACTAAATGCAAACTT----- 1152
 Db 790 PheValSerAspAspLysLeuThrAspValTyrThrLysMetSerAlaGluValThrAsn 809
 QY 1152 ----- 1152
 Db 810 AlaGluGlyIleLysLysGluIleAlaGluLysGluPheGluAsnValHisLysLysMet 829
 QY 1153 -----TTTGAAGTTTGGTTAGGAAATCAACA--- 1179
 Db 830 LysGluPheSerAspAlaPheSerThrLysPheGluAlaLeuGluAsnSerMetGluGln 849
 QY 1180 -----GACACATTTTAAACAAACAAACAAACAAATATGAAACAAATACAA 1224
 Db 850 TyrAsnGluGluGluLysAspAlaIleGluLysHisLysGluAsnArgSerGluLysGluGln 869
 QY 1225 TCATATTTCGACACATATTCACAAATTTGCAATATTAATATGTAATTTTAAACAA 1284
 Db 870 GluTyrPheLysAsnGluSerValGluGluAspLeuSerArgLysGluThrGluGluGln 889
 QY 1285 TTTTATGAAAACCTTAG----- 1302
 Db 890 GluTyrThrLysHisLysAsnAsnPheSerArgArgLysGluLysIleSerAlaGluIle 909
 QY 1303 -----GAAAGCGCA-----TATGCACTAAT 1323
 Db 910 ThrAsnMetArgGluValIleAsnLysIleGluSerGluLeuAsnTyrTyrGlyValIle 929
 QY 1324 GACACTTTTAAATTACTCA-----AATGAAGAAAGATATGTAAGAGAGATTA 1374
 Db 930 GluLysTyrPheSerLeuIleGlyAspGluAsnGluValSerThrAlaLys----- 946
 QY 1375 CCAGACGAAAGATATTTACTTTTACTACAGCTGATGACAAAGATATTTATCGT 1434
 Db 947 ---AlaLeuLysGluLysIleValSerAspSerLeuArgAspLysIleAspGluTyrGlu 965
 QY 1435 TCAGAAATAT-----TGCCAACTGTCTCCGACTGC 1464
 Db 966 ThrGluPheSerGluLysThrSerAlaValGluAsnThrValSerThrIleGlnSerLeu 985
 QY 1465 GGGGCTCAATGCGATGCTATTAATATACACACAAATTCACATATGAT----- 1512
 Db 986 SerLysAlaIleAspSerLeuLysArgLeuAsnGlySerIleAsnAsnLysLysLysTyr 1005
 QY 1513 -----CGTGAACGTGTAATAAT 1530
 Db 1006 AsnThrAspLLeuAspLeuArgSerLysIleLysThrLeuArgGluGluValGluLys 1025
 QY 1531 GAAGACTATAA-----CCTCCATGGGGTGAGACCTACTAATATCTACTCTTTAT 1584
 Db 1026 GluMetProLysArgLysAspLysCysGlyGluAsnThrThrAlaLeuLeuLysSer 1045
 QY 1585 AGTGTAAATGACAAAGCTGATATTACACAAAAATTTGAAAAATTTTGTACACTCACT 1644
 Db 1046 LeuArgAspLysMetGlyLysIleAsnGluLysLeuAsnAspLysArgLeuAsnSerLeu 1065
 QY 1645 AATTACAAAGATATAATATCAAAATGGGATGCTATTA----- 1686
 Db 1066 AspThrLysLysGluLysLeuLys-----PheTyrSerGluSerLysSerLys 1082

QY 1687 -----AAGATGAAAT-----ATAATATGATGTAACCTGA 1719
 Db 1083 IleHisLeuSerLysAspGluLysGlyProGluAspProLeuAsnLys-----IleAsp 1100
 QY 1720 CAAATATACGAATTCATTAATGATATCCATACATA-----ATATCATTTCTTAATTTTTT 1776
 Db 1101 GluTyrGluLysPheLysArgAspValAspGluLeuAsnValAsnTyrGluValIleSer 1120
 QY 1777 GAATTATGGGTACTCATATTATTAAGGACTATTAAGGAAATGCAAACTTAAC 1836
 Db 1121 GluAsnLysValThrLeuPheLysAsnAsnSerValThrTyrIleGluAlaMetHisSer 1140
 QY 1837 TGTATAAATATACCAACACGACTGTATTTGATGAATGTAAACAAATTCCTATGCTTT 1896
 Db 1141 HisIle---AsnThrValAlaHisGlyIleThrSer---AsnLysAsnGluIleLeu--- 1157
 QY 1897 GACAGATGGTTAAACAAAGAAAGAAATGCAATATGATTAAGAAACCTTCACAAA 1956
 Db 1158 ---LysSerValLysGluValGluAspLysLeuAsnLeuValGlu----- 1171
 QY 1957 AAAAAGATATACGCAATGCTATTAATATGTAATTAATAT----- 1998
 Db 1172 -----GluAsnGluAspTyrLysLysValLysAsnProGluAsnGluLysGln 1187
 QY 1999 -----CTTTTGAAGTTATTTTAAAGTATGATGATTAACCTGACAAAGAT----- 2046
 Db 1188 LeuGluAlaIleArgLysSerMetSerLysLeuLysGluValIleAsnLysHisValSer 1207
 QY 2047 -----GACGAAATGGAAGAACTTATGCAAAATATTAAGAAACAA 2091
 Db 1208 GluMetThrGluLeuGluSerThrAlaAsnThrLysSerAsnAlaLysGlyLysGlu 1227
 QY 2092 AATGATGTTTCAATTTGCAAAATATAGGACTATTTAGACATGCAATGCAATGCACTG 2151
 Db 1228 AsnGluHisAspLeuGluGluLeuAsnLys-----ThrLysGlyGlnMetArgAspIle 1245
 QY 2152 TTGATATCCTTAAAGAACTGCCAGATATGTAAAGACAAATATTAACAAAGCAAGATGT 2211
 Db 1246 TyrGluLysLeuLysLysIleAlaGluLysGluLeuLysGluThrValAsnGluLeuLys 1265
 QY 2212 GAAACATCCCATATATGCAACAAACAAACCCGTGTAAACCTGTGAGGCACGCAACCC 2271
 Db 1266 AspAlaAsnGluLysAlaAsnLys-----ValGluPro-----GluPro 1278
 QY 2272 ACTTAAATATTAAGCAATATGCAATATTAAGAGTGCATATGCAAGAGACGA 2331
 Db 1279 GluArgAsnIle-----IleGlyHisValLeuGluArgIleThrValGluLysAspLys 1296
 QY 2332 AAT-----CGTGGTCTTCATTAATGAAAGAAAGCAGCAGAGGTATATAT 2379
 Db 1297 AlaGlyLysValValGluGluMetAsnSerLeuLysThrLysIleGluLysLeuIleGln 1316
 QY 2380 AAACGTGGGGGTAGGAAAGCACTTCAAGACAAATTAATGTAATATGATTAATGATTA 2436
 Db 1317 GluThrSer-----AspAspSerGluAsnGluLeuValThrThrSerIleThrLys 1333
 QY 2437 CATTTCAATGCTATCTGTTTTCATTAATGACATGATGCAAAAGCAGCAGATGAT 2496
 Db 1334 HisLeuGluAsnAlaLysGlyTyr----- 1341
 QY 2497 GGTATACAAACAAAGATTTGCTAGAACATGGAAGTGGAACTGCAGAACACATGCGT 2556
 Db 1342 -----GluAspValIle 1345
 QY 2557 AAAGATCAGCAAGATGTTATTAATGCTCCAGAAAGACGATATATGTAATGCAATTTG 2616
 Db 1346 LysArgAsnGluLysAspSerIleGlnLeuArgGluLysAlaLysSerLeuGluThrLeu 1365
 QY 2617 GAACATTTACAAACGATGATCACCCACTTAATGTAATGTTGTTGATGATTAGTTAAT 2676
 Db 1366 AspGluMetLysLysLeuValGluGlnValAsnMetAsnLeuGlnSerAlaIleGlnGly 1385

QY 2677 AATTCCTTTTGGGGAGTCTCTATCAGCAAAATATGAAACAAGATATATAGCA 2736
 Db 1386 AsnAlaGlyLeuSer----- 1390
 QY 2737 ATGTTAAAGAAATACCTAAAGGCCCCAAAGATGACTGACCAAAACACGAC 2796
 Db 1391 -----LysGluLeuAsnGluLeuLysGluValIleGluLeuLeuIleSerThrAsnTyr 1408
 QY 2797 ACAACTATCTGTCGACGATACGTTTGGCATATAGGTGATATATATTCGAGA 2856
 Db 1409 SerSerIle-----LeuGluTyrValLysLysAsnSerSerGluSerValAlaRphe 1425
 QY 2857 AGAGTCTCTGGGAAAGAAACGGTGACATGGTAAAGCTGCAAGGACATTTGGAAACTGTT 2916
 Db 1426 SerGluLeu-----AlaAsnGlyGluPheThrLysAlaGluGluVal----- 1439
 QY 2917 TTTGGTATATATACATTAAGTCTACTCAAAAGCAAGAAATGATATATATGATGATGCC 2976
 Db 1439 ----- 1439
 QY 2977 CCCAAATTTTAAATGAGGAAAAATTTGGGAGACTAAT--AGAGCCAAAGTATGG 3033
 Db 1440 -----GluLysAsnAlaSerAlaAlaRgLeuAla 1448
 QY 3034 GAAGCCATGAATGTGATATATAATATTTGAGGATTAATCGGACACCAATCAACAA 3093
 Db 1449 GluAlaGluLysLeuLysGluGluIleValLysAsp----- 1460
 QY 3094 AGTAGTATTGCGGATATAGTATCATACCACTTGATGATTTATTCGCCAAATAATTA 3153
 Db 1461 -----LeuAspTyrSerAspIleAspAspLysValLys----- 1471
 QY 3154 AGATGATGACCGAATGGCAGAAATGCTACTGCAAGGTGACAG--AAAAAGGCTAT 3207
 Db 1472 -----LysIleGluLysIleLysAlaRgLeuIle 1480
 QY 3208 GATTAAGTTGAGAGAGAGTGTAAAGAGTAAAGATTAAGATTAAGTCAAGCTGTACG 3267
 Db 1481 LeuLysMetLysGluSerAla----- 1487
 QY 3268 AAGAAGATGTGTACAGTGTGACAGAGTGCACAGAACTTTGAATGAATATATATATATA 3327
 Db 1487 ----- 1487
 QY 3328 ATAGGATTTTGAAGAAACAATGAATATATATATGATATATATATATATATATATGA 3387
 Db 1488 LeuThrPheTyrGluGlu-----SerGluLysPheLysGluMet----- 1500
 QY 3388 CAAGCAACAATGTCTGTAGTATAGTGTATGTAAGCTTCCAGTACGCAAAAATCAT 3447
 Db 1501 -----CysSerSerHisMetGluAsnAlaLysGluLysLys 1514
 QY 3448 ATAGCAGAGATGTATTGAATTTTGTGCGAATTAATACCAACAATGTTGGGCAAAAGT 3507
 Db 1515 IleGlu-----TyrLeuLysAsnAsnGluAspGly 1524
 QY 3508 AATAAAGTGTACTAGTAT----- 3528
 Db 1525 GLyysAlaAsnIleThrAspSerGluMetGluValGluAsnTyrValSerLysAla 1544
 QY 3529 -----GAAAGTCTCTCATTTGTTGTTACTTAACACCAAGTATGAAT 3567
 Db 1545 GluHisAlaRheHisThrValGluValAlaGluValAspLysThrLysAlaRheCysGluSer 1564
 QY 3568 GTTGAGCATATCTCCATGATACAGGAATTTTATGATATGTCATGACCAAAATGAGTTT 3627
 Db 1565 IleValAlaTyrValThrLysMetAspAsnLeuPheAsnGluSerLeuMetLysGluVal 1584
 QY 3628 -----TGTGATGAAAAAAGTATGATGAAGTAAACGAAAAATATGCTTTAGAGAT 3678
 Db 1585 LysValLysCysGluLysLysAsnAsp-----GluAlaGluLysTyrSerAlaLysLeu 1602
 QY 3679 AAACCAACAGACCATGATGCTGCTGTGTTGAAAAAGTATGAAACGCAAGAGGTA 3738

Db 1603 LysPro-----TyrAspGlyAlaGlyIleLysAlaRgValSerGluAsnGluAlaRgIle 1620
 QY 3739 ---CAGTTAAAGAAAGAAAAAGCGGAAGAAAGATACGAAATGTAAACAGTAAAT 3795
 Db 1621 SerGluLeuLysGluLysAlaLysValGluLysLys-----GluSerSerGluLeuAsn 1638
 QY 3796 GATATACCTTAAAGAAAAACGATGGAAGAAACAAGTGAAGATTCATCCAAAAAGAAAT 3855
 Db 1639 AspVal-----SerThrLysSerLeuGluIleAspAsnCysAlaRgGluIleLeuAsp 1656
 QY 3856 AGTAATGATATCCGATTCGCAATGCGGAATATTAATTTAGTGAAGACCCCTGTCG 3915
 Db 1657 Ser-----ValLeuSerAsnIleGlyAlaRgVal----- 1665
 QY 3916 TGTATCCCCCTAGAACCAAAAGTATTCGCAATTTCTTGTGCAATGATATGAATA 3975
 Db 1666 -----LysIleAsnAlaLeuGluIleTyrPheAspSerAlaAspLysSerMet 1680
 QY 3976 AAAAAATTACAA-----TCACAAGTTAATTTAAAGAAAGCTTTTCATCAAAATCTGCACGA 4029
 Db 1681 LysSerValLeuProIleSerGluLeuGluAlaGluLysSerLeuAspLysValLysAla 1700
 QY 4030 CGAGAAACATTCCTCATGATATATATATAAGTAAAGATGATGGAAGAAATGAATC 4089
 Db 1701 AlaLysGlu-----SerTyrGluLysAsnLeuGluThrValGluAsnGluMet 1716
 QY 4090 GATTA--GAATTAAAGAAAGCAAAATTCCTCCGCAATTTTGAGATCCATGTCATC 4146
 Db 1717 SerAlaGluLeuAsnValGluGluGlySerLeu----- 1726
 QY 4147 ACATTTGAGATTTATAGATTTTATTTTATTTGACATATATCAAAAGT----- 4197
 Db 1727 -----ThrAspIleAspLysLysIleThrAsp 1735
 QY 4198 ---CATGCTAGGAGAGTAAACTAAAGACGA-----ATAGATTCCT 4239
 Db 1736 IleGluAsnAspLeuLeuLysMetLysGluLeuIleTyrGluGluLysLeuGluIle 1755
 QY 4240 TTCAAAAATGCTGACCAAAAATCCTTAATGGAAGAAACACGCCAAGAAATGTGACAGAA 4299
 Db 1756 LysGluAsnAlaAspLysValG-----LysSerAsnPheGluLeuValGlySer 1771
 QY 4300 CATATGATGATATATGGAAGCAAGTATGCTATGCTACATTAATAATTTGGGCAAAAAA 4359
 Db 1772 GluIleAsnAlaLeuLeuAspProSerThrSerIlePheIleLysLeuLysGlu 1791
 QY 4360 GATGATTTTACG-----GAAACTACGCT----- 4383
 Db 1792 TyrAspMetThrGluAspLeuLysLysAsnTyrGlyValLysMetAsnGluIleIleGlyGlu 1811
 QY 4384 -----TACAAACAAGTCAAA-----TTTATGCACAAAAGCAC 4416
 Db 1812 PheThrLysSerTyrAsnLeuIleGluThrHisLeuSerAsnAlaThrAspTyrSerVal 1831
 QY 4417 ACTTTGAGCAATTTGCCAAAGCAAGCCAGCTTTTACGATGCTTAACCAAGTGAAGAC 4476
 Db 1832 ThrPheGlu-----LysAlaGluSerLeuThrGluLeuAlaGluLysGluGlu 1847
 QY 4477 GACTATTTCTATACACAGCA----- 4497
 Db 1848 GluHisLeuAlaGluGluGluGluAlaIlePheLeuLeuAsnAspIleLysLysVal 1867
 QY 4498 -----AATATTTTGAAGATGTCGAGGAATAATGTAAGTCAAT----- 4536
 Db 1868 GluSerLeuLysLeuLeuLysGluMetLysLysValSerAlaGluTyrGluGluMet 1887
 QY 4537 -----GACCA 4542
 Db 1888 LysAlaGluPheHisThrSerValSerGluLeuValGluAspMetLysThrIleValAspIle 1907
 QY 4543 TTGAAGTGT-----GATACAGAAATGTATTAAGAAATGCGAGACTAGTAAATAT 4593

D 1908 LeuLysThrLeuAsnAspIleSerGluCysSerSerValLeuAsnValIleSerIle 1927
Q 4594 ATGAAAAAAGAGAGTGGATTCACAAAGATAATATTACAGATGACCGCAAA 4653
D 1928 ValIleValSerValIleSerValIleSerValIleSerValIleSerValIle 1943
Q 4654 AAAAGATTGATGACAAACATTGGTGAATGTTACACATATCTGGAAGATGCA 4713
D 1944 AsnSerMetCysGluSerMetValIleSerValIleSerValIleSerValIle 1954
Q 4714 ACAGATTACTTGAACAAATTTACTGCTAGTTGGTGATAGCCGTGGAAGTGC 4770
D 1955 AsnTyrPheLeuSerAspGluAlaIleSerSerGlyMetGluPheAsnAlaGluMet 1974
Q 4771 -----TCTGTGACCAAGAATATA 4791
D 1975 LysSerAsnPheLysThrAspLeuGluIlePheSerValIleSerAsnSerAsn 1994
Q 4792 CAATTGTTGAAAACAGCGTTCTACTGATGCCGACAAACATGTGGGTGCACAAATTT 4851
D 1995 GluLeuLeuLysLys 1999
Q 4852 ATTGAAATGACGCAAAATTTACTAATTTTCGATTAAGATAGTCAAGATAGTA 4911
D 2000 IleGluGlnAspSerAsnAspValIleGluLysGluArgGluSerGluGlnLeuAla 2018
Q 4912 AAGGAGCCAAACACAGCTGCTAATAGTGCAAAACAGCTCTAATTAACAATAC 4971
D 2019 LysAspAlaThr 2026
Q 4972 TTGAAAGAAATGACTGAAGATGTGCTTTCTCTGCTGACCTAGCTAATGTTTCAT 5031
D 2026 ----- 2026
Q 5032 GCATTGATGCAATTATACAGATCCAGAAAGTAAAGTAAAGTGGTTCGCAAAAGA 5091
D 2027 -----ValIleLysLeuLysAsnGluPheAsnGluLys 2037
Q 5092 TTGATGGAAGTGGCGGACAGGAGGTACAAATTTGGGTCAATACACAAAGAAAAA 5151
D 2038 LeuGluGluAla 2045
Q 5152 GAAAAAGAAAAATAAACGTGCGATGCGCACAAATATTTCTATGAGTCCGCCCTGT 5211
D 2046 GluValValSerGluLysValArgGluAlaLeuLysArgLeuSerGluValIleGluGlyIle 2065
Q 5212 AGTCTATGAATATGATTTTATGATTTAAGATATTAATTTAGCTATGATTAATTTG 5271
D 2066 ArgCysHisPheGluAsnPheHisArgLeuLeuAspAsnThrGluGluLeuAsnLeu 2085
Q 5272 GAA-----GATGAAAAACAAAGACCGAAGAAAAATTTGAACAA 5310
D 2086 LysLysMetValThrIleTyrArgAspLysLysSerGluArgGluSerGluGlnGlu 2105
Q 5311 ATTTTAACAAAAATGGAATCATGTTGGCAAGAAAGATAGTACTAGAGAAATCCC 5370
D 2106 MetGluAsnGluMetAsnThrTyr 2121
Q 5371 GGTAGTACTCCGCAAAATTTTCTGAAACGAAATTAAGCAATGCTGGAACGCAATG 5430
D 2122 GlyIleValValSerAla 2134
Q 5431 ATATGCGCGGTACAAACGTGTAGGATGATGAATAGTGGAATAGTGCACAGAGTGA 5490
D 2135 -----GluLysLeuGluArgSerAsn 2141
Q 5491 GAAGATCTAAAA-----AAATGTTGTTCTGTACTCTCA----- 5523
D 2142 GluGluMetArgAsnIleSerGluLysIleSerThrIleAspSerLysValIleGluMet 2161
Q 5524 -----GATGATGATTAATCTTANGGAAAAAATCGCATGAAGGCTACTGCTAT 5571
D 2162 AsnSerThrIleAspGluLeuTyrLysLeuGlyLysAsnGlyAsnGln 2176

Q 5572 CAGTTTCTTGATGCTTGGCCGATGGGTGAAGATTTTGCACAACTAAAGAAAGAA 5631
D 2177 -----AlaHisThrIleSerLeuIleSerTyrThrAlaAsnMetLys 2190
Q 5632 TTGGAGAAATTTGGAGGGCGGTGAATGATTATTAATCTGTGTATATGAAGATAAAGA 5691
D 2191 ThrSerLysLysLeuIleMetIleAsn 2201
Q 5692 AGAAATGTACAGATGCGGTGACACAAATATATAATTTATAGTGAAGGAAACACAG 5751
D 2202 LysGluAsnThrGluLysCysValAspTyrIleLysAspAsnSerSerThrAspGly 2221
Q 5752 TATGAAAAACAAATCAAAAAATTTGTGACAAATTAAGACAAATATATTCGAGCATCT 5811
D 2222 TyrValGluThrLeuLysGlyPheTyrGlySerLysLeuThrPheSerSerAlaSerGlu 2241
Q 5812 GTGCAAAAGATGACAGG-----GAGCGCTGCAATATTTAGCAAAACATTA 5859
D 2242 IleValGlnAsnAlaAspThrTyrSerValAsnPheAlaLysHisGluLysGluSerLeu 2261
Q 5860 AAAAAATTTGTGAAAAATAAAGTGAAGTGTGAATATATAGTGAAGATGTGTC 5919
D 2262 AsnAlaIleArgAspIleLysLysGluLeuTyrLeuPheHisGlnAsnSerAspIleSer 2281
Q 5920 ACACAGCATTAACGTATGTAATAGTCAAAATATGCCCGCA-----TCATTAGAC 5970
D 2282 -----IleValGluGlyGlyValGlnAsnMetLeuAlaLeuTyrAspLysLeuAsn 2298
Q 5971 GATGAAACCAAAAGAGTGAAGAAAGTGAATATGTCAATGGCACAGAGTCCACACAGT 6030
D 2299 GluGluLysArgGluMetAsp 2306
Q 6031 GTACGAAGGGAACACCGGTACACAGCGGTATCACTGATATCAAAACGACGCGATCGAA 6090
D 2307 LeuTyrArgAsnIleSerGluThrIleLysLeuLysGluMetGluHisSerThrAspValPhe 2326
Q 6091 AAAGAGCGAAAAACGCGCCCTACAAACACGCGGAAAAAGTGGAAATCTTAACACA 6150
D 2327 Lys-----PrometIleGluLeuHisLysGlyMetAsn 2337
Q 6151 GAAATGCGACACAAACAGCACGCCGACGACGACACACAAACACAAACGAAACATCA 6210
D 2338 -----GluThrAsnAsnLysSerLeuGluLysGluLysLysLysSer 2353
Q 6211 ACAGCAACAACAACAGATCTGACGTGGCCACATGTAAGGCCATCTTTCGAATAAA 6270
D 2354 ValAsnAspHisMetHisSerMetGluAlaGluMetIleLysAsnGlyLeuLysTyrThr 2373
Q 6271 CCAGATAGCAGGGGTGGAATAGAG----- 6294
D 2374 ProGluSerValGlnAsnIleAsnAsnIleTyrSerValIleGluAlaGluValLysThr 2393
Q 6295 -----GCTTGAATCCAAAAACGATGACAAATCTCTAA 6330
D 2394 LeuGluGluIleAspArgAspTyrGlyAspAsnTyrGluIleValGluGluHisLysLys 2413
Q 6331 -----TGGGTTGTATTGTAGCTAGCTC----- 6354
D 2414 GlnPheSerIleLeuIleAspArgThrAsnAlaLeuMetAspAspIleGluIlePheLys 2433
Q 6355 AAAGAAAAATGAAAAATGCAATATGTATGCTCTTAGAGAAAAAATTAATATTAATAT 6414
D 2434 LysGluAsnAsn 2439
Q 6415 ATCAATATTTAATTTAAGCTGAAGTAAAGCGTACATGATATAAAGAGGCTTT 6474
D 2440 LeuMetGluValAsnThrGluThrIleHisArgValAsnAspTyrIleGluLysIleThr 2459
Q 6475 ATTAATGTGCAACCAATAGAACTCAATTTTGTGTTAAATATATATTAATGAATCCT 6534
D 2460 AsnLysLeuValGlnAlaLysThrGluTyr-----GluGlnIleLeuGluValAsnIle 2476

```

OY 6535 GCACACAAAAATGAAATTCGAAATGCAACATTCAGATGATTTAAAGATATATGAT 6594
Db 2477 LysGlnAsnAspMetLeuGlnAsnIlePheLeuLysValSerIleIleGluTyr 2496
OY 6595 TATACATATGATGATTAAGATATGTTTGGACACGATATTTCTAATGATAAAAA 6654
Db 2497 PheGlnAsnValLysLysLysGluSerIleLeuAsnAspLeuTyrGluGlnGluTyr 2516
OY 6655 ATTTATACGTCAACAAATAGTACACCATTT-----CTCAATGAAAAATATAG 6705
Db 2517 LeuLeuLysIleGluLysIleAsnGluIleLysArgAsnValThrGluThrLeuSer 2536
OY 6706 AAAAAACAGCATAAAAAAGATGAGATTAACGTAATATTTGGAGAAAAATATAA 6765
Db 2537 SerTyrGluLysAspGlnMetSerLysAsnLeuLeuGluLysLysSer 2556
OY 6766 AATTTATTTGGAGAGATATATGATGATTAATGATTCATCTACACAGCAAAAC 6822
Db 2557 LysMetMetAsnTyrThrSerIleTyrGluLeuGluValAsnGluLysAsnArg 2576
OY 6823 GAAAAAGAAAAATAGATATATACGATACATGACATGACCAACAGCGCTTCC 6882
Db 2577 AspAlaLysGlnIleLysAsp-----AspAspThrIleLeuAsnSerVal 2591
OY 6883 CTTCGAGAGCTTGTGA--AAAAAGCCCAATTTTGAGATGCTGCACAGAAATGC 6933
Db 2592 LeuGluAlaIleIleGlnLysArgLysPheLysAlaIlePheSerGlnMetSerAla 2611
OY 6934 -----GCAGAGAAATTTGTATATATAGAGAGAGAGAGAGAG 6966
Db 2612 AspArgAsnProAsnGluTyrLysSerAlaGluLysTyrMetAsnGluAlaAsnGluIle 2631
OY 6967 TTGTTAAATTTGAGCGCGCTGTAGAGATATAGTATAGTATAGTATAGCGGTAG 7026
Db 2632 IleArgGlnLeuGluValLysLeuArgGluIle-----GlyGln 2644
OY 7027 ACACAGAGATGTGAGAGCGCTGTACACATACATCAAAATTTATTAAGAGTGAAG 7086
Db 2645 LeuValGlnAspSerGluSerIleLeu-----Ser 2654
OY 7087 GAATATGAAAGACAAAGAGAAAGTTCAAAAAGATTAAGATGCAAAAAGCTTAAGAT 7146
Db 2655 GluMetAsnSerLysLysSerAlaIleGluLysGluLysThrAlaArgAlaLeuArgThr 2674
OY 7147 TATCCTCTCTGAGAGAGACATAGAGAGCAACATGCTCATGAAATTTAAACATG 7206
Db 2675 SerGluAsnAsnArgArgGluGluGluIle-----ArgAla 2686
OY 7207 AAATTAAGAAATTAATGTGCAATAGATTTGTTGTATGCAAAAACCTTCTACAA 7266
Db 2687 ArgValGlnGluMetSerMetAsnAsnAspProThr-----GlnSerGluThrThrHis 2704
OY 7267 CTTCGAAAAACACACACATCACAATCCGATCGATGATGATGATCCGATCGCTG 7326
Db 2705 SerGluGlySerIleGluLysGluSerAspSerAspGluThrGluLeuThrHis 2724
OY 7327 GATTATGTTCTGAGAGATTAATTAACAAGTGTGCTGCAACTTTGAAAAAGGATCT 7386
Db 2725 AspAlaGluValAspGluAspSerThrSer-----SerAlaLysGluVala 2739
OY 7387 ATGATTCATACAAAAAATTAAGTGAACCTTAATACCTATGAATTTGTGAGAAAGCA 7446
Db 2740 -----HisGluLeuGlnGluGluThrAlaProMet----- 2751
OY 7447 GCATATATTTATTAAGAGAGCAAAATATATGATATTTACCTTG----- 7494
Db 2752 -----GluGluThrGluMetAsn--AspAsnThrLeuLeuGluTyrAsp 2765
OY 7495 -----AAGAAAAATTTATACCTATGATGCTACAAAGAA-----AAG 7533
Db 2766 ThrThrArgSerAspGluProAspMetHisThrGluAsnThrGlnAspGlyThrTyrGln 2785
OY 7534 GAAAGTAAAAATAGTGGAGCTAATAATATATTCCTTGGATCTTAAGAAACCTTATGACCT 7593

```

```

Db 2786 AspThrSerAsnSerAspGluAlaAspIleLeuAsnGlyLysPheAsnProAsn 2805
OY 7594 GATTAATATATAGCA 7608
Db 2806 ValLysTyrAlaGly 2810

RESULT 3
NEBU_HUMAN
ID NEBU_HUMAN STANDARD; PRT; 6669 AA.
DC P20929; O15346;
DT 01-FEB-1991 (Rel. 17, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nebulin.
GN NEB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95257391; PubMed=7739042;
RA Labelt S., Kolmerer B.;
RT "The complete primary structure of human nebulin and its correlation
RL to muscle structure.";
RN [2]
RP PARTIAL PRELIMINARY SEQUENCE.
RX MEDLINE=88284704; PubMed=3397062;
RA Bevilani M., Darras B.T., Rizzuto R., Salvietti G., Betto R.,
RA Bonilla E., Miranda A.F., Du J., Samlitz C., Dickson G., Walsh F.S.,
RA Dimauro S., Francke U., Schon E.A.;
RT "Cloning and expression of human nebulin cDNAs and assignment of the
RL gene to chromosome 2q31-q32.";
RN [3]
RX Genomics 2:249-256(1998).
RN [4]
RP STRUCTURE BY NMR OF 6610-6669.
RX MEDLINE=98179559; PubMed=9514727;
RA Pollou A.S., Millevol S., Gautel M., Kolmerer B., Pastore A.;
RT "SH3 in muscles: solution structure of the SH3 domain from nebulin.";
RN [5]
RX J. Mol. Biol. 276:189-202(1998).
CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
CC THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
CC ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
CC OF STRIATED MUSCLE.
CC -!- DISEASE: Defects in NEB are a cause of the autosomal recessive
CC form of nemaline myopathy (NEM2).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 178 NEBULIN REPEATS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; X83957; CAA58788.1; -
DR EMBL; M19668; AAS59916.1; ALT_SEQ.
DR EMBL; M19669; AAS59917.1; ALT_SEQ.
DR PIR; A29979; A29979.
DR PIR; B29979; B29979.
DR PDB; IARK; 28-JAN-98.
DR PDB; IARK; 24-DEC-97.
DR Genew; HGNC:7720; NEB.
DR MIM; 161650; -.
DR MIM; 256030; -.
DR InterPro; IPR000900; Nebulin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.

```

DR pfam: PF00880; Nebulin, 146.
 DR PRINTS; PRO0452; SH3DOMAIN.
 DR SMART; SM00227; NEBU; 181.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0002; SH3; 1.
 KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
 FT REPEAT 176 107 NEBULIN 1.
 FT REPEAT 113 143 NEBULIN 2.
 FT REPEAT 148 178 NEBULIN 3.
 FT REPEAT 183 213 NEBULIN 4.
 FT REPEAT 218 248 NEBULIN 5.
 FT REPEAT 253 283 NEBULIN 6.
 FT REPEAT 289 318 NEBULIN 7.
 FT REPEAT 324 354 NEBULIN 8.
 FT REPEAT 363 393 NEBULIN 9.
 FT REPEAT 398 428 NEBULIN 10.
 FT REPEAT 434 464 NEBULIN 11.
 FT REPEAT 502 532 NEBULIN 12.
 FT REPEAT 537 567 NEBULIN 13.
 FT REPEAT 573 603 NEBULIN 14.
 FT REPEAT 611 641 NEBULIN 15.
 FT REPEAT 681 711 NEBULIN 16.
 FT REPEAT 749 779 NEBULIN 17.
 FT REPEAT 820 850 NEBULIN 18.
 FT REPEAT 858 888 NEBULIN 19.
 FT REPEAT 893 923 NEBULIN 20.
 FT REPEAT 924 954 NEBULIN 21.
 FT REPEAT 959 990 NEBULIN 22.
 FT REPEAT 993 1023 NEBULIN 23.
 FT REPEAT 1028 1058 NEBULIN 24.
 FT REPEAT 1064 1094 NEBULIN 25.
 FT REPEAT 1102 1132 NEBULIN 26.
 FT REPEAT 1137 1167 NEBULIN 27.
 FT REPEAT 1168 1198 NEBULIN 28.
 FT REPEAT 1204 1234 NEBULIN 29.
 FT REPEAT 1237 1267 NEBULIN 30.
 FT REPEAT 1272 1302 NEBULIN 31.
 FT REPEAT 1308 1338 NEBULIN 32.
 FT REPEAT 1346 1376 NEBULIN 33.
 FT REPEAT 1381 1411 NEBULIN 34.
 FT REPEAT 1412 1442 NEBULIN 35.
 FT REPEAT 1448 1478 NEBULIN 36.
 FT REPEAT 1481 1511 NEBULIN 37.
 FT REPEAT 1516 1546 NEBULIN 38.
 FT REPEAT 1552 1582 NEBULIN 39.
 FT REPEAT 1590 1620 NEBULIN 40.
 FT REPEAT 1625 1655 NEBULIN 41.
 FT REPEAT 1656 1686 NEBULIN 42.
 FT REPEAT 1692 1722 NEBULIN 43.
 FT REPEAT 1725 1755 NEBULIN 44.
 FT REPEAT 1760 1790 NEBULIN 45.
 FT REPEAT 1796 1826 NEBULIN 46.
 FT REPEAT 1834 1864 NEBULIN 47.
 FT REPEAT 1869 1899 NEBULIN 48.
 FT REPEAT 1900 1930 NEBULIN 49.
 FT REPEAT 1936 1966 NEBULIN 50.
 FT REPEAT 1969 1999 NEBULIN 51.
 FT REPEAT 2004 2034 NEBULIN 52.
 FT REPEAT 2040 2070 NEBULIN 53.
 FT REPEAT 2078 2108 NEBULIN 54.
 FT REPEAT 2113 2143 NEBULIN 55.
 FT REPEAT 2144 2174 NEBULIN 56.
 FT REPEAT 2180 2210 NEBULIN 57.
 FT REPEAT 2213 2243 NEBULIN 58.
 FT REPEAT 2248 2278 NEBULIN 59.
 FT REPEAT 2284 2314 NEBULIN 60.
 FT REPEAT 2322 2352 NEBULIN 61.
 FT REPEAT 2357 2387 NEBULIN 62.
 FT REPEAT 2388 2418 NEBULIN 63.
 FT REPEAT 2423 2453 NEBULIN 64.
 FT REPEAT 2456 2486 NEBULIN 65.
 FT REPEAT 2456 2486 NEBULIN 66.

FT REPEAT 2491 2521 NEBULIN 67.
 FT REPEAT 2527 2557 NEBULIN 68.
 FT REPEAT 2565 2595 NEBULIN 69.
 FT REPEAT 2600 2630 NEBULIN 70.
 FT REPEAT 2631 2661 NEBULIN 71.
 FT REPEAT 2666 2696 NEBULIN 72.
 FT REPEAT 2699 2729 NEBULIN 73.
 FT REPEAT 2734 2764 NEBULIN 74.
 FT REPEAT 2770 2800 NEBULIN 75.
 FT REPEAT 2808 2838 NEBULIN 76.
 FT REPEAT 2843 2873 NEBULIN 77.
 FT REPEAT 2874 2904 NEBULIN 78.
 FT REPEAT 2909 2939 NEBULIN 79.
 FT REPEAT 2942 2972 NEBULIN 80.
 FT REPEAT 2977 3007 NEBULIN 81.
 FT REPEAT 3013 3043 NEBULIN 82.
 FT REPEAT 3051 3081 NEBULIN 83.
 FT REPEAT 3086 3116 NEBULIN 84.
 FT REPEAT 3117 3147 NEBULIN 85.
 FT REPEAT 3152 3182 NEBULIN 86.
 FT REPEAT 3185 3215 NEBULIN 87.
 FT REPEAT 3220 3250 NEBULIN 88.
 FT REPEAT 3256 3286 NEBULIN 89.
 FT REPEAT 3294 3324 NEBULIN 90.
 FT REPEAT 3329 3359 NEBULIN 91.
 FT REPEAT 3360 3390 NEBULIN 92.
 FT REPEAT 3395 3425 NEBULIN 93.
 FT REPEAT 3428 3458 NEBULIN 94.
 FT REPEAT 3463 3493 NEBULIN 95.
 FT REPEAT 3499 3529 NEBULIN 96.
 FT REPEAT 3537 3567 NEBULIN 97.
 FT REPEAT 3572 3602 NEBULIN 98.
 FT REPEAT 3603 3633 NEBULIN 99.
 FT REPEAT 3638 3668 NEBULIN 100.
 FT REPEAT 3671 3701 NEBULIN 101.
 FT REPEAT 3706 3736 NEBULIN 102.
 FT REPEAT 3742 3772 NEBULIN 103.
 FT REPEAT 3780 3810 NEBULIN 104.
 FT REPEAT 3815 3845 NEBULIN 105.
 FT REPEAT 3846 3876 NEBULIN 106.
 FT REPEAT 3914 3944 NEBULIN 107.
 FT REPEAT 3949 3979 NEBULIN 108.
 FT REPEAT 3984 4014 NEBULIN 109.
 FT REPEAT 4021 4052 NEBULIN 110.
 FT REPEAT 4057 4087 NEBULIN 111.
 FT REPEAT 4088 4118 NEBULIN 112.
 FT REPEAT 4123 4153 NEBULIN 113.
 FT REPEAT 4156 4186 NEBULIN 114.
 FT REPEAT 4191 4220 NEBULIN 115.
 FT REPEAT 4226 4256 NEBULIN 116.
 FT REPEAT 4264 4294 NEBULIN 117.
 FT REPEAT 4299 4329 NEBULIN 118.
 FT REPEAT 4330 4360 NEBULIN 119.
 FT REPEAT 4365 4395 NEBULIN 120.
 FT REPEAT 4400 4430 NEBULIN 121.
 FT REPEAT 4435 4465 NEBULIN 122.
 FT REPEAT 4471 4501 NEBULIN 123.
 FT REPEAT 4544 4574 NEBULIN 124.
 FT REPEAT 4575 4605 NEBULIN 125.
 FT REPEAT 4610 4640 NEBULIN 126.
 FT REPEAT 4645 4675 NEBULIN 127.
 FT REPEAT 4680 4710 NEBULIN 128.
 FT REPEAT 4716 4746 NEBULIN 129.

Alignment Scores:

Pred. No.: 8, 41e-13
 Score: 420.50
 Percent Similarity: 30.948
 Best Local Similarity: 17.838
 Query Match: 2.16%
 DB: 1
 Gaps: 197

US-10-087-013-1 (1-10628) x NEBU_HUMAN (1-6669)

QY	81	TATATAAAGTGCACAAAGCGCAG--AAATGTTT-----GGAACGTATTGCCAA	131
Db	705	TYTYSALAGLUTYGLUGLUSPRLYSCTYTRPHEPROGLTHRLIEHRGln	724
QY	132	AAATTAAGACATCCATCAAAATA-----TGCAAAAGACATGTGGATGCGTGAAGG	185
Db	725	GLUTYASPAIAlIeLYLSLEUSPRLInS-lysaspHsthrTYLysValHisPr	744
QY	186	GGATTTGCGAAGACGAATTTCCGTGGTGCCTTCAAGCCAGTAATAAGCATATTA	245
Db	744	oAspLysThrLys-----PheThrAlaValThrAspSerProVal-----	757
QY	246	TTAATTCATATCCATGTATTAATTAGATCATAGAGACATCTAATTACGGTAT----	300
Db	758	-----LeuLeuGlnAlaGlnLeuAsnThrLysGlnLeuSerAspLeuAsnTYLysAl	775
QY	301	-GATGATGTGAAATTGGAGACATCTTGCCAT-----	330
Db	775	ALYSHisGLUGLUGLUTRPhelyScysHisLeProAlaAspAlaProGlnPheIleGl	795
Db	331	-----GGTAGAAGAACAAACCGAATTGCATGAACATGAAGATTCGA	371
Db	795	nHisArgValAsnAlaTYrAsnLeuSerAspAsnValTYLysGlnHisPrTrgLUlys--	814
QY	372	ATGTGGAATTAATACGTATTAATTAAGAAAGAAATGATGCTTACCTGTGGCCACC	431
Db	815	-----SerLysAlaLysLysPheAspRIleLysValAspAlaIleProleuAlaAl	832
QY	432	TAGAAGACGACATATGTGTATTAATAACCTTGGAAGCTCTAAATGATTAATCCCAAA	491
Db	832	ALys-----	838
QY	492	TATTCATGATTTATTTGGAAATGTACTAGTTACAGCAAAATACGAA-----	537
Db	838	nThrSerAsp-----ValMetTYrLysLysAspTYrGluLysSerLysGly	854
QY	538	-----GGGAAATTCATTTTAAATATCATCCACATTAAGAACTTCAGACCTGTGAC	590
Db	854	SmetIleGlyAlaLeuSerIleAsnAspAspProLys-----	866
QY	591	TGCTCTGCACGAAGTTTGGACATATAGTGATATTGTGAAGAGAAATAGATNTGTTAA	650
Db	866	-----	866
QY	651	ACCAATGTCCATGACAAAGTAGAAACGGCTCCGACAGGTTTTCAGAAATATACATGA	710
Db	867	-----MetLeuHisSerLeuIleThrAlaLys	876
QY	711	TGGAATGGAAGATGAGTAAATAATGATTATACATCCTGATGGATTCGAAATATATATA	770
Db	876	nGlnSerAspAlaGluTYrArgLysAspTYr-----GluLysSerLysThrIleTYrTh	894
QY	771	ATTTAAGACAGCATGGTGAATGTGAATAGAAATAAAGTATGAGATATACATGTA	830
Db	894	rAlaProLeuAspMetLeuGlnValThrGlnAlaLysLysSerGlnAlaIleAlaSerS	914
QY	831	TGCATCATTAATATCTGGATTTTATGCAATCAGAACTAATACACCATTAATTTTCAA	890
Db	914	pValAspTYrLys-----HisIleLeuHisSerTYrSerTYrProProAspSerIleS	932
QY	891	TCCTAATGCGCGCCATAAACAAGAAAGGTTCCTACCAATTAAGATTATGTCCTCATTA	950
Db	932	nValAspLeuAlaLysLysAlaTYrAlaLeuGlnSerAspValGluTYrLysAlaAspTY	952
QY	951	TTTACGTTGGTTC-----	963
Db	952	rAsnSerTrpMetLysGlyCysGlyTYrValProPheGlySerLeuGluMetGluLysAl	972
QY	964	-----GACGAATGGCGGACGAGAGAGTTTCCCGCAAAAGAAATATTAAATGAA	1010
Db	972	ALysArgLAsSerAspIleLeuAsnGluLysLysTYrArgGlnHisProAspThrLeuLY	992
QY	1011	A-----AAGCGCAAGACCTCCTGCG	1031
Db	992	sPheThrSerIleGluAspAlaProIleThrValGlnSerLysIleAsnGlnAlaGlnAr	1012
QY	1032	TAATGAC-----	1038
Db	1012	gSerAlleAlaTYrLysAlaLysGlyGluGluIleIleHisAsnTYrAsnLeuProPr	1032
QY	1039	-----AAAGACGGCTTAATTTGATGCATTAATAGACATATGTCAGCACTATTG	1091
Db	1032	oAspLeuProGlnPheIleGlnAlaLysValAsnAlaTYrAsnIleSerGluAsnMetY	1052
QY	1092	GAAAAAGATTTTGGCATTTTGGATTAATAGTCTACTGACGTCCGACTAAATGCAAGT	1151
Db	1052	rLysAlaAspLeuLysAspLeuSerLysGlyTYrAspLeuArgThrAspAlaIlePr	1072
QY	1152	TTTTGAGTTTGGTTAGGAAATCAACAAGACGA-----TTTAAAAACAAA	1199
Db	1072	oIleArgAlaAlaLysAlaAlaAlaAlaAlaAlaAlaSerAspValGlnTYrLysAspTY	1092
QY	1200	AGAAAAATATGAAAAAGAAATATACATCATAT-----TTATGCAAGCATTAACAAATTTGT	1253
Db	1092	rGluLysAlaLysGlyLysMetValGlyPheGlnSerLeuGlnAspProLysLeuVa	1112
QY	1254	CAATTAATTAAT-----AGTAAATATTATTAACAATTTATGAAAA	1295
Db	1112	IHisTYrMetAsnValAlaLysIleGlnSerAspArgGluTYrLysLysAspTYrGluY	1132
QY	1296	ACTTAAAGCAACCAATATGCACTATATGACACTTTTAAATTTACTAATGAAGAAA	1355
Db	1132	sThrLys-----SerLysTYrAsnThrProHisAspMetPheAsnValAlaAlaLysLY	1151
QY	1356	GTTTGTGAAGAGATTTACAGAGAAAAAGATATTACTTTTACAAACAGCTCGATGA	1415
Db	1151	s-----AlaGlnAspVal-----ValSerAsnValAsnTY	1161
QY	1416	CAAAAGGATTTATTCGTTTCAGAAATATGCGCAAGTGTGCCGATCGGGGTCAAAGT	1475
Db	1161	rLysHisSerLeuHisHisTYrThrTYrLeuProAspAlaMetAspLeuGluLeuSerLY	1181
QY	1476	TGATGGTATTAATATACACACACAAATCATATGATGCTGGAACGTGAATATGAAGA	1535
Db	1181	sAsnMetMetGlnIle-----GlnSerAspAsn-----ValTYrLysGluS	1195
QY	1536	CTATTAACCTCCATCG-----GGTGAAGCCTACTAATATGACSTGTCCCTTA	1583
Db	1195	pTYrAsn-----AsnTrpMetLysGlyIleGlyTYrIleProIleGlySerLeuAspAlaGl	1214
QY	1584	TAGTGTATTAAGCAAGGTGATATT-----	1610
Db	1214	uLysValLysLysAlaGlyAspAlaLeuAsnGluLysLysTYrArgGlnHisProAspTh	1234
QY	1611	ACAAAAATTAGAAAAATTTTGAACAGCTCAACTAATTAACAAGATTAATAATTAACAAA	1670
Db	1234	rLeuLysPheThrSerIleValAspSerProValMetValGlnAlaLysGlnAsnThrLY	1254
QY	1671	ATGC-----GAATCCTATTATTAAGAGGAATAATTAATAGATGTAACCTGGAACAAA	1724
Db	1254	sGlnValAspAspIleLeuTYrLysAlaLysGlyLysP-----ValLysHisLY	1271
QY	1725	TACTGAAATCATTAATGATTAATCCTAGATATATATCATTTTCAAT-----AATTTTGGATTT	1781

```
Db 1323 -----Serat 1324
OY 1962 GAATATACAGCAATCGATATAGTATATATTAATCTTTTGAAGGTATTTTAA 2021
Db 1324 gsnlllealaseraptyr-----1330
OY 2022 AGTTATGATTAACCTTGACAAAGATGAGCAAAATGCAAAAGACTTATGCAAAATATAA 2081
Db 1331 -----LysTyrLysGluAlaTyrGluLysSerLys 1340
OY 2082 AAAAAAATGAGTT-----2100
Db 1340 sglyshsvalgllypheatrgserleuglnasparprolylsleuvalhistrymetas 1360
OY 2101 -----TCAATTTGGAATAATATAGGACATTTAGAAATGCAATAGAACTGTGTTAGA 2156
Db 1360 nvalalalysleuglnaseraprglululyslysasn-----TyrGl 1375
OY 2157 TCACCTTAAAGAAAGCCGACGATATGTAAGCAATTAATACAAAGCAAGCATGTGAAC 2216
OY 1375 uasntrlysthrserTyrhistryrproglyasrmetvalthrlethralalalsme 1395
OY 2217 ATGCCATTAATGCAAAACCCGCTGTCTTAACCTCGTGAGCGACGCAACC-----2271
Db 1395 talaglnasprvalalathrasnvalasnTyrLys-----GlnProleuH1 1410
OY 2272 -----ACTAAATATATAAAGAAAT 2291
Db 1410 shistrythrTyrLeuProaspralameTserleuglnhistryrghsnvalasnGlnI1 1430
OY 2292 AGCACAATACTTTAAAGAGTCATACGAGAAAGCAAGT-----CGTGCTC 2342
Db 1430 eglIn-----SerAsprsnvalTyrLysAsprglulTyrAsnsrpherleuysglYI1 1447
OY 2343 T-----CATTAATTAAGAAAGCACACGAGG 2372
Db 1447 eglTTPrlleproileglyserleuglnvalglulysvalys-----LysAlaeglYasprAl 1466
OY 2373 TATATTAACGGGGGGTGGAGA-----AAGCACTTCAAGGA 2411
Db 1466 aleuasnglularglulysTyrarglnhistryrproasprthvalylsrpherhistryrvalProas 1486
OY 2412 CAATTTATGAGATTAATGATTAACCAATCTTAATCGTAATCTTGTTTCAATGAGCC 2471
Db 1486 psermetglYmetleuvalaglnhistryrproasprthvalylsrpherhistryrvalProas 1506
OY 2472 ATGTGANTGCAAAAGCCAGCTGATGTATACAAACAAGTTTGTCTAGCACTGATG 2531
Db 1506 svalgluglYglu-----LysleuLysshistry 1516
OY 2532 GGAAGTGCATCCGAGACATGCGTAAGATCAGAAAGATGTTATGTTCCTCTAGAG 2591
Db 1516 rthrllleasprproglu-----leuPro-----1523
OY 2592 ACACATATATGATCATCAATTTGCAACATTTACAAACGAGATCACCCCACTTAATGC 2651
Db 1524 -----Glnpheileglnalalysvalasnlaalysn-----1534
OY 2652 TAATATGTTGATGATTACTTAATAATCTTTGGGGAGTGTCTTCTACAGCAAA 2711
Db 1535 -----metSerAsprAlaH1 1539
OY 2712 ATATGAGCAAAAC-----AAGATATACGAATGTATTAAGAAAGAAATACCTTAAGCG 2765
Db 1539 styrlYslalasprrlysthrlysthrlearg-----LysGlyTyrtrspreleuargProas 1557
OY 2766 CCCCAGAAAGTACGACCCCAACACCCAGACACTATGTCTGACCTTACGTTACAG 2825
Db 1557 palalileproilevalalalalysserSerArgasnillealaseraprcyslystYrly 1577
OY 2826 TTTTGCATATAGCTGATTAATTCGAGAAAGATCTCTGGAAAGAAAGCGTGACAT 2885
Db 1577 sglnAla-----1579
OY 2886 GGTAAAGCTCAGAGACATTTGGAAACTGTTTGGTAAATATACATAAGTCAAGG 2945
Db 1580 -----TyrGluYslalalsgl 1585
OY 2946 CAAA-----GGAATGATTAATATATGATGATGCCCCCAATATTTAAATTTGAAGCGA 2999
Db 1585 YLysGlnvalglYpheatrgserleuglnaspar-----ProlylsleuValhistryr-----1602
OY 3000 AATTTGTTGGGAAGCTTAATAGACCAAGATATGGAAGCCATGAATGTATTAATA 3059
Db 1603 -----Metasnvalalalsyle-----GlnSerAsprglulY 1614
OY 3060 TTGAAGATTAATCGGACCAATCAACAAAGATGATTTGCGGATATGATGATCA 3119
Db 1614 T-----LysLysGlyTyrGlnalaserlysthrlystYr-----H1 1626
OY 3120 TACACCATTTGGAT-----3132
Db 1626 sthrProleuAsprmetvalserValthrAlalalysSerGlnGluvalAlaThrAs 1646
OY 3133 -----GATTATATCCCAAAAAATTAAGA-----3156
Db 1646 nalaasnTyrArgGlnserTyrhistryrthrleuLeuProasprAlaleuasnvalGl 1666
OY 3157 -----TGATGACGCAATGCGCAGA 3176
Db 1666 uhlsserArgasnAlameTlnlleGlnSerAsprasnleuTyrLysSerAsprhistryrAs 1686
OY 3177 ATGC-----TACTGCAAGTGCAGAAAAAGCACTATGATTAAGTGAAGA 3221
Db 1686 nTPrmetlysglyleuglyrprvalProileGlnserleuglnvalglulysvalalsyl 1706
OY 3222 GAAGTGTAAAGAGCTTAAGATTAAGATTAATGCTCAAGCTGTACGAAAGAGTGTAC 3281
Db 1706 salaglyGlnleuSerleuLysTyrArgGlnhistryrproglululysleuysrpheth 1726
OY 3282 AGCTGTACCAAGTGCACACAGCTTGT-----AATGAATATGATGATTAAT 3329
Db 1726 rTyrAlameTasprhmetleuglnalaleuasnlyserhistryrvalalsyl 1746
OY 3330 AGCATTAATGGAAGAACATGCAAT-----3354
Db 1746 savgleuTyrThrGlnulysstrasnlysnlysthrThrillehistryrvalmetProasprTh 1766
OY 3354 -----3354
Db 1766 rProasprlleuLeuSerArgvalasnGlnllethmetSerAsprlysluTyrLysAl 1786
OY 3354 -----3354
Db 1786 aglyTTPrlleuglnulyslysglyTyrAsprleuArgProasprAlalealilely 1806
OY 3355 -----ATAATATCAGAT-----AATTAACAAGATTAATACATGACA 3389
Db 1806 salalalalargalaserArgasprillealaserAsprTyrlysthrlyslalarglul 1826
OY 3390 AGCACAATGCTGTGATAGTATAGTGTATGAAGCTCCAGTACCCAAATATAT 3449
Db 1826 nalaLys-----GlyLysshistry 1832
OY 3450 A-----GACAGAAATGTTATTAATTTTGT-----3474
Db 1832 eglYpheatrgserleuglnasparprolylsleuValhistryrmetcullvalAlalysme 1852
OY 3475 -----TCGAATTAATACCAAAATGCTGCCAAAGATTAATTAAGTGT-----3519
Db 1852 tGlnSerAsprglululyslysglyTyrGlnulysSerlysthrserpherhistryrpr 1872
OY 3520 -----ACTAGTGAAGAAAGTGTGCTATGCTAGTACAGACACAC 3557
Db 1872 ovalasprmetleuSerValvalalalalyslysserGlnGluvalAlaThrAsnAlas 1892
```



```

OY 7134 AAGATTAAGAT-----TATCCTTCTACTGAAAGAGA 7166
Db 3140 eTYrLYsSerAspLeuGlnTrpLeuArgLYleGlyTrpValProIleGlySerMetAs 3160
OY 7167 CATAGAGAGGCAACATGCTGCTCATGAAATTTAAAC----- 7203
Db 3160 pValValLYsCysLYsArgAlaAlaGluIleLeuSerAspAsnIleTYrArgGlnProPr 3180
OY 7204 -----ATGAAATTTAAAGAAATTTATGTGCAATAAGA 7235
Db 3180 oAspLYsLeuLYsPheThrSerValIThrAspSerLeuGlnValLeuAlaLYsAsnAs 3200
OY 7236 TTGTCTTGTATGCAAAAACCTTCTCACACAGTACCAAAAACAACACAAATCACAAATC 7295
Db 3200 pAlaLeuAsnMetAsnLYs-----ArgLeuTYrThrGluAlaIleTrpAspLYsAspLY 3217
OY 7296 ATCCGATGCTAAATGATATGCGCAATGCGGATTTATGCTCCGAGAAATTTAAACAAGTG 7355
Db 3217 sThrGlnValHisIleMetTrpAspThrProGluIleMetLeuAlaArgGlnAsnLYsII 3237
OY 7356 TGAGTGTCTGAACCTTCAAAAAAGGAGTCTATGATTCATACAAAAA----- 7404
Db 3237 eAsnTYrSerGlnSerLeuTYrArgGlnAlaMetGlnGluAlaLYsGlnIleTYrAs 3257
OY 7405 -ATTACTGAACCTAAATACCTATGAAATGTGTAGAGAAA-----GCACCAATA 7451
Db 3257 pLeuAlaGSerAspAlaIleProIleValAlaAlaLYsAlaSerArgAspIleAlaSerAs 3277
OY 7452 TTTATTTATTAAGAGAGCAAAATTAATATG----- 7482
Db 3277 pTYrLYsTYrLYsGluAlaTYrArgLYsGlnLeuGlnHisIleGlyAlaArgAlaVa 3297
OY 7483 -----GATATTAC 7490
Db 3297 IHisAspAspProLYsIleMetTrpSerLeuHisIleAlaLYsValGlnSerAspArgI 3317
OY 7491 CTTGAAGAAAAATTT-----ATACCTATGTAGAGCTTACAAA 7526
Db 3317 uTYrLYsLYsAspPheGlnLYsTYrLYsThrArgTYrSerSerProValAspMetLeuG 3337-
OY 7527 GCAAAAAGAAATTAATAGTTGCACTAATTAATATCTTCCGATCCTTAAGAAACCTTA 7586
Db 3337 YlleValLeuAlaLYsLYsCysGlnThrLeuValSerAspValAspTYrLYsHisProLe 3357
OY 7587 T-----GCACCTGATAAA-----TATATNGAAGAAAGAAACCTTGGA 7625
Db 3357 uHisGlnCysIleCysLeuProAspGlnAsnAspIleIleHisAlaArgLYsAlaTYrAs 3377
OY 7626 AATAGAGAGAAAAATCTTTTAAGGTAGATTATGAAATGAAATGTTACAAAAATTCAAA 7685
Db 3377 pLeuGlnSerAspAsnLeuTYrLYsSerAspLeuGlnTrpMet----- 3391
OY 7686 GTTCTATCAGAGAAAAAGATATGTGTACCTCCAAAGAGAAACATATGTGCTTAAG 7745
Db 3392 -----LYSGLYIleGlyTrpValPro----- 3398
OY 7746 GAATTAAGATGAATTAATTAATGAAGACTTAAGATAGTAATTTCTCTAAATAATGTT 7805
Db 3399 ----IleAspSerLeuGlnValValArgAlaLYsArgAlaGlyGlnLeuLeuSerAspTh 3417
OY 7806 TCGTGAACTGCACGAATGAAGAGATAGACATTAATAAAAACTTCAACTCA----- 7857
Db 3417 rIleTYrArgGlnArgProGlnThrLeu-----LYSPheThrSerIleThrAs 3433
OY 7858 -----GAGAACGGGTGCGCAATGCAATATGATATGATATGAAATATAGTTTGGC 7910
Db 3433 pThrProGlnGlnValLeuAlaLYsAsnAsnAlaLeuAsnMetAsnLYsArgLeuTYrTh 3453
OY 7911 TGATCTGGGTGATAGTATTAAGAGAAACAGATATGTTACGAATGGTGGTACTTACCTCC 7970
Db 3453 rGlnAlaIleTrpAsp-----AsnAspLYsLYsThrIleHisValMetProAspTh 3469
OY 7971 CGTAAAGATTAATATATAGGTTTGTGATATACATATATGAAATGAGAAATTAATAA 8030
Db 3469 rProGlnIleMetLeuAlaLYsLeu----- 3477
OY 8031 TAAAGTAGAAATTAATATACAGAT-----GTACAAAGCTTTCGTTTC 8072
Db 3478 ----AsnArgGlyLeuAsnTYrSerAspLYsLeuTYrLYsLeuAlaLeuGlnGlnSerLYsLY 3496
OY 8073 TCGTTGTGGATGCTTAATAGAAAGATTTTGAAGCAATGACGTGCAAAGCAACGA 8132
Db 3496 sGlnGlyTYrAspLeu-----ArgLeuAspAlaIleProIleGlnAlaAlaLYsAlaSerAr 3515
OY 8133 AGATGCAAAACCTTTTAGAAGAGAAAGATGATGATTTGAACGATTAACATTAATACA 8192
Db 3515 gAspIleAlaSerAspTYrLYsTYrLYs-----GlnGlyTYrAlaGlys----- 3529
OY 8193 AGATAAGTGTGACAT-----AAGACGATCCACCTGTTGATGA 8231
Db 3530 ----GlnLeuGlnHisIleGlyAlaArgAsnIleLYsAspAspProLYsMet----- 3546
OY 8232 TTAATACCTCAACGTTTCGATGATGACGTGATGCTGAAATATTTGTAAGCACT 8291
Db 3547 -----MetTrpSerIleHisValAlaLYs----- 3554
OY 8292 GATGAGAAATTTGAATAATTTAAATATCATGTGATCAGTGAATACATGTCGACAGATG 8351
Db 3555 -IleGlnSerAspArgGlnTYrLYsGlnPheGlnLeuLYsTrpLYsThr-----LYSPH 3572
OY 8352 CAAGAATGATTTATGATGAATAATAGTGCACAGTGTAAACGAGATGTCAA----- 8403
Db 3572 eSerSerProValAspMetLeuGlnValValLeuAlaLYs-----LYSCysGlnIleLeuVa 3591
OY 8404 -----GAAATTAATAATTTGTTCTTAATATGAAATCCTCATGAT----- 8445
Db 3591 lSerAspIleAspTYrLYsHisAspProLeuHisGlnIleTrpThrCysLeuProAspGlnAsnAs 3611
OY 8446 ----ATACATCAATTAATATCAAGAATTTGTATGACAAACCAATATATACAAAAATCTC 8501
Db 3611 pValIleGlnAlaArgLYsAlaTYrAspLeuGlnSerAspAlaIleTYrLYs-----Se 3629
OY 8502 TACTATGATCATGTTCAAAATTT-----GTACAAA 8534
Db 3629 rAspLeuGlnIleTrpLeuAlaGlyIleGlyTrpValProIleGlySerValGlnValGlnLY 3649
OY 8535 GTTGAACCTTTTAAAGTAGATGTTCTGTGAGAGCTTTCGTGA----- 8580
Db 3649 sValLYsArgAlaGlyGlnIleLeuSerAspArgLYsTYrArgGlnProAlaAspGlnLe 3669
OY 8581 -----TATCTCATGAGAAACAGTATAGTGTGAAATTAATTAATGA 8630
Db 3669 uLYsPheThrCysIleThrAspThrProGlnIleValLeuAlaLYsAsnAlaLeuTh 3689
OY 8631 TGGTCTTCCAAATATACAGATATGCTTCGAA----- 8664
Db 3689 rMetSerLYsHisLeuTYrThrGlnAlaIleTrpAspAlaAspLYsThSerIleHisValMe 3709
OY 8665 -----AAAAGTTATTAAGAACTTGCAGTTGTACACT 8705
Db 3709 tProAspThrProAspIleLeuLeuAlaLYsSerAsnSerAlaAsnIleSerGlnLYsLe 3729
OY 8706 ACTTCTAAGATCATGATGATATATGCTACCGAGCAAAACAAAGATGAGATGAAGA 8765
Db 3729 uTYrThrLYsGly-----TrpAspGlySerLYs----- 3738
OY 8766 ATTAACAACCTTTTACCTTGTGCTGGAAGATATGATATATATATCTGATATTTGAA 8825
Db 3739 -----MetLYsAspTYrAspLeuAlaAlaAsp----- 3747
OY 8826 CGCATACCTGTCTTAATAGTCAAGTGAATGATTAACAAGAGGTATTTGATTTCTTCAAGAG 8885
Db 3748 ----AlaIleSerIleLYsSerAlaLYsAlaSerArgAspIleAlaSerAspTYrLYsTY 3766
OY 8886 AAGACATTTATGTACAAGACCTATCACTGATTAATTAATTAAGAAAGGT----- 8934

```

Dd	4046	rargleuhsprtrialals-----		4052
Oy	9903	gttagcgtttgaaagccgttcatttcatgaagaamaaatccaatgcstcgactctgt		9962
Dd	4053	-----trpylsalaysileginserrprovalasmethe		4064
Oy	9963	ggctactacgaatatccccgaagagactgatgaaatgccctacgttgaaatccaaaatag		100222
Dd	4064	userlleuhns-----serlysaense		4072
Oy	10023	gtacatfaccatpatagaaagtgcctcaatpaaggcamaacatpatatat-----		100711
Dd	4072	rglnalaueualaserarmelasttyr-----argsanitytleuhisglntprlthcy		4090
Oy	10072	-atggaagagactactagtcgagatgaagatpaataatpatgpcgsaacattatctcctgca		101308
Dd	4090	smeethroasprclnabnasrvaalllelnlalysylsalatyrgluuenglinserspas		4110
Oy	10131	tattacttatccgamaagtactat-----gaagaaatt		101632
Dd	4110	nvalitrylsalaserleuglutprleuargilyleglytrmetproasnsrvaser		4130
Oy	10164	cgattttaat-----gatpatatgaccagsgatgctctaattatpa		102055
Dd	4130	lseralasnhsialalyshslalalaspliehe-----serglusylasytyatr		4148
Oy	10206	aacattgatagaaagtactgtaacaacatcaaaaaagcattatccaaagcatgatracacc		102658
Dd	4148	gthrlyslleglutprleu-----asnherthrrprovalaspsarlayva		4163
Oy	10266	aagtaatagtatcacccacct--acsaatratgatatttgatgat-----ga		103077
Dd	4163	lasptyvalthratalayrlnglnsersegglyglutlleuaspasrilelystyraglysas		4183
Oy	10308	atggaatgacagcaaatcatgatattt-----		103327
Dd	4183	ptrasnalathrlxrselystythrleuthrglutprroleuhenhsithralagl		4203
Oy	10333	-----gtaatccaattatttaccmaatraacagaacamaataatgaattacaa		103767
Dd	4203	ngluhalaalatrgtileudasperglutytleu-----tytly		4215
Oy	10377	aagtcagatnatccaaatgaanaacgaacstaattattatcttgatnaatccsgaaga		104363
Dd	4215	sgluely-----trgluatrglunalsalathrglytleuportospaal		4232
Oy	10437	aaaaccttttatattatcatgacatvtagcatttatatctggcaaaatpatagtaa		104968
Dd	4232	avalprobevalnhsalnhsishscysasmrvalginser-----gluleulysty		4250
Oy	10497	T 10497		
Dd	4250	r 4250		
RESULT 4				
ID	S230_PLAFO	STANDARD:	PRT: 3135 AA.	
AC	008372:			
DT	01-FEB-1995 (Rel. 31, Created)			
DT	01-FEB-1995 (Rel. 31, Last sequence update)			
DT	15-JUL-1999 (Rel. 38, Last annotation update)			
DE	Transmission-blocking target antigen S230 precursor.			
GN	S230			
OS	Plasmodium falciparum (isolate NF54), and			
OS	Plasmodium falciparum (isolate 3D7).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
OX	NCBI_TaxID=5843, 36329;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN=Isolate NF54;			
RX	MEDLINE=93241227; PubMed=8479460;			
RT	Williamson K.C., Criscio M.D., Kaslow D.C.;			
RA	"Cloning and expression of the gene for Plasmodium falciparum			

QY 1864 ATTGATGAATGTAAACAGAAATTCCTTATGTTTGGACAGATGCGGTTAAACAAAAGAGACA 1922
 :::::||||| ||||| ||| :::::|||||
 Db 105 PhgIUmEcGlySerLysAsnIleThrcysPheTyrrProIleValGlyLysGluArg 124
 1924 GAATGAATAGT-----ATAAGAACTGTTTCCAAAAAAGAAATATACAGCA 1974
 LysThrLeuAspThrIleIleIleLysAsnValThrAsnAspHisValIleSer 144
 1975 TCGATTATAGTAATATATATATATCTTTTGAAGCTATTTTAAAGTATATGATTA 203
 :::::|||||:::
 Db 145 AspMetHisSerAsnVal----- 150
 QY 2035 CTTCGACAAAGATGAAAGCAAAATGGAAAAACTTATGCAAAATATA---AAAAAGAAAAA 2091
 ::::: ||| ||::: ||||| |||
 Db 151 -----GlnGluLysAsnMetIleuIleIleArgsnIleAspLysGluAsnLys 166
 QY 2092 AATGAGTTTCCAAATTGGAAAAAT-----AATAGGACATATTAGACAAATGCAATAGCA 2145
 ||||| ||||| ||||| |||||
 Db 167 AsnAspIleGlnAsnValGlnGluLysIleGlnIleAspThrTyrrGluAsnLys----- 184
 QY 2146 CTCCTGTAGATCACTTAAAGAAATCCACAGATA---TGTAAAGACATATATACAAAC 2202
 ||||| ||||| ||| ||||| |||||
 Db 185 -----AspTyrrGluSerAspThrLeuIleGluThrPheAspAsnThrAsn 201
 QY 2203 GAAGATGTGAACATCCCATATATGCAACAACAAACCCGTGTGTTAAACCTGCGGAAGC 2262
 ||| ||| ||| ::::: ||| ||| |||
 Db 202 Glu-----GluAsnPhe-LeuLeuThrPheLeuLysArgCysLeuMetLysIlePheSe 219
 QY 2263 ACCGAACCCACTAAAAATATATAAAGAAATAGACACATACTTAAAGAGCGATACACAG 2322
 ||| ||| ||| ||||| |||||
 Db 219 rserProLysArgLys-----LysThrValValGlnLys 230
 QY 2323 GAAGACAGAAATCGTGTCTTCATTAATTGAAAGAAAGCAGACAGAGATATATATA 2382
 |||||::: ||||| |||||
 Db 230 sLysHisLysSerAsnPheIleAsnSerSerLeuLysTyrrIleTyrrMetLysLeuTh 250
 QY 2383 CGTGGGGGTAGAGAAAGAGACTTCAAGGACAAATTTGTAGAAATATATATAAACATTTCT 2442
 r-----ProserAspSer-PheAsnLeuValArg-----Arg 260
 250 r-----ProserAspSer-PheAsnLeuValArg-----Arg 260
 QY 2443 AATCGTAATCTTGGTTTTCAAAT---GGACCATGTGATGGCAAGCAGGTGATGT 2499
 ||||| ||||| ||| ||| |||
 Db 261 AsnArgAsnLeuAspGlnGluAspMetSerProArgAsp----- 273
 QY 2500 ATACAACACAGATTTGTCGTAGAGACTGAAATGGAAAGTGGATCCGAGACATCGCTAAA 2559
 ||||| ||||| ||| ||| |||
 Db 274 -----AsnPheValIleAspAspGlnGluGlnGluGlnGluGlnGluGln 290
 QY 2560 GATCAGCAAGATGTTATTTATGCTCTCTAGAAAGCAGCATATGTATCAATTTGAA 2619
 ::: ||||: ||||| |||||
 Db 291 GluGlnGluGln-----GluGlnGlnGln 298
 QY 2620 CATTTACAACGATGATGATCCACCACTTAATGCTAATATTGTTGATGATTAAGTAAATAT 2679
 ::::: ||||: ||||| |||||
 Db 299 GluGlnGlnGlnGlnGlnGlnGln-----AspAspTyrrVal----- 309
 QY 2680 TCCTTTGGGGGATGTCTCTATCAGCAAAATATGAAGCAACAGATATATCGAATG 2739
 309 ----- 309
 QY 2740 TATTAAGAAAGATTAACCTAAAGGGCCCAAGAAAGTAACTGACCCAAACACAGACACA 2799
 ||||: ||||: ||||| ||||| |||||
 Db 310 TyrrGlnGlnGlnGlnGlnGln-----GluThrGlnGlnGlnGlnGlnGlnGlnGln 327
 QY 2800 ACATATC---TGTGACGCTATAGCTTACAGTTTGTGAGATATAGGT---GATATATTTCCA 2853
 ::: ||| ||||| ||| ||| |||
 Db 328 GluValGlnLysLysSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 347
 QY 2854 GGAAGATCTCTGGGAAAGAAACGATGATGATGATGATGATGATGATGATGATGATGATGAT 2913
 ||||| ||||| ||||| ||||| |||||
 Db 348 AAlArgAsp-----GlyAspMetIleArgValAspGlnLysTyrrGlnLys 362

Db 914 ---AsnAsnLysGluHisAspTyrThrCysAspPheThrAspLysLeuAspLysThrVal 932
QY 4876 AACATTTGAGTAAGTAAG-----TGCAAGGATTAAGTAAGGAGCGCAAC 4923
Db 933 ProSerThrIleAsnGluLysLysLeuPheIleCysArgLysHisLeuLysGluPheAsp 952
QY 4924 ACAGGTCCTTAAGTGGCAACAAAGGTCCTAATACATACATTAATGAAAGATG 4983
Db 953 ThrPheThrLeuLysCys---AsnValAsnLysThrGlnTyrProAsnIleGlu----- 969
QY 4984 ACTGAGATGCGTTTTCCTCCGTCGACATACGTAATGTTTTCATGCAATGGATGCC 5043
Db 970 -----IlePheProLys----- 973
QY 5044 AATTATACAGATCCAGATTAAGATGAATAAGGTTGCGAAAGATGATGAGATG 5103
Db 974 -----ThrLeuLysAspLysLysGluValLeuLysLeuAspLeuAspIle 988
QY 5104 -----CGGCGAACGGAAGGATCAAT 5124
Db 989 GlnTyrGlnMetPheSerLysPhePheLysPheAsnThrGlnAsnIleLysTyrLeuAsn 1008
QY 5125 TTGGGTCAATACTACAA-----GAAAAAAGAAAAAGAAAA 5163
Db 1009 LeuTyrProTyrTyrLeuIlePheProPheAsnHisIleGlyLysLysGluLeuLysAsn 1028
QY 5164 ATAAAAACGTCGATGCCCAATATTTTATGAGTCCCGCTTGATGCTATGAA 5223
Db 1029 AsnProThrTyrLysAsnHisLys-----Asp 1037
QY 5224 TATAGTTTATGATTAAAGATTAATCTAGTATGATTAATTTGAGAT----- 5277
Db 1038 ValLysTyrPheGlnIleSerValIleSerProLeuSerSerIleAspSerLeuGly 1057
QY 5278 -----GAAAAACAAAGACGAGAAAAATTTGAGAAAAATATTAAAC 5319
Db 1058 LysLeuLeuAsnPheLeuAspThrGlnGluThrValCysLeuThrGluLysIleArgTyr 1077
QY 5320 AAAAAATGCAACATCTATGGCAAGAAAGTATGATCTACACAGAAATCCCGTAGTACT 5379
Db 1078 LeuAsnLeuSerIleAsnGluLeuGlySerAspAsnAsnThrPheSerValThrPheGln 1097
QY 5380 GCGCGAAATTTTCTGGAAGCAAAATAGAAATGCTGTGAGACCATGATATGCGGG 5439
Db 1098 ValProTyrIle-----AspIleLysGluProPheTyrPheMetPheGlyCysAsn 1115
QY 5440 TACAAACGTGTGATGATGAAATAGTGA-----AATAGTCAAGAACT 5487
Db 1116 AsnAsnLysGly-----GluGlyAsnIleGlyIleValGluLeuLeuIleSerLysGln 1133
QY 5488 GATGAAGATCTAAAAAATGCTGTTCTGACTTACATGATGATGATATTCCTATGGGAAA 5547
Db 1134 GluGluLysIleLysGlyCysAsnPheHisGluSerLysLeuAspTyr----- 1149
QY 5548 AATCCGATGAAGTACTGCGTATCATGTTCTTCGATGGTTGCCAAATGGGTGAAGAT 5607
Db 1149 ----- 1149
QY 5608 TTTTGCACAATAAAGAAAGAAATTTGAGAAATTTGATGAGGCGCTGTATATCATTTACT 5667
Db 1150 -----PheAsnGluAsnIle 1154
QY 5668 TGTGCTGAATTAAGATTAAGAAAGAAATGTACAGATGCGGTGTACACATATATAAAA 5727
Db 1155 SerSerAsp-----ThrHisGluCysThrLeuHis----- 1164
QY 5728 TTTATTAGTAGTGAACCCAGATGATAAAACAAATCAAAATATATGATGAGAAATAA 5787
Db 1165 -----AlaTyrGluAsnAspIleIleGlyPhe----- 1173
QY 5788 GACAAATATATTCGACATCT-----CTGCGAAAGATGACAGACGCTCGCGAA 5841
Db 1174 AsnCysLeuGluTyrThrHisProAsnGluValGluValGluValGluAspAlaGluIle 1193
QY 5842 TATTGACAAACATTAATAAAAAATTTGTGAATAAAGTGAAGTGTGAATTAAG 5901
Db 1194 TyrLeuGlnProIle-----AsnCysPheAsnAsn-----ValTyrLys 1206
QY 5902 TGTATGAAGATGTGTCACACAGCATTAATGATGATGAT-----AAT 5943
Db 1207 GlyLeuAsnSerValAspIleThrThrIleLeuLysAsnIleGlnThrTyrAsnIleAsn 1226
QY 5944 AGTCAAAATATGCGCCGATCATTAAGACATGAACA-----AAAGAGTTGAA 5991
Db 1227 AsnLysLysThrProThrPheLeuLysIleProProTyrAsnLeuGluAspValGlu 1246
QY 5992 GGAAGTGAATTTGTCATGATGCCACAGAGTCCACACAGTGA----- 6033
Db 1247 IleSerCysGlnCysThrIleLysGlnValValLysLysIleLysValIleIleThrLys 1266
QY 6034 -----CGAAGGGAACACCGTCA----- 6051
Db 1267 AsnAspThrValLeuLeuLysArgGluValGlnSerGluSerThrLeuAspAspLysIle 1286
QY 6052 -----CCACGGTATCAGTATA----- 6069
Db 1287 TyrLysCysGluHisGluAsnPheIleAsnProArgValAsnLysThrPheAspGluAsn 1306
QY 6069 ----- 6069
Db 1307 ValGluTyrThrCysAsnIleLysIleGluAsnPheAsnTyrIleGlnIlePheCys 1326
QY 6070 ---TCAAAAGCAGCGCATCGAAAAAGAGAGAAA-----ACAGCGCGGCT 6114
Db 1327 ProAlaLysAspLeuGlyIleTyrLysAsnIleGlnMetTyrTyrAspIleValLysPro 1346
QY 6115 ACAAAACGCCG-----AAAAAGTGAAT-----CTACA 6147
Db 1347 ThrArgValProGlnPheLysLysPheAsnAsnGlnGluLeuHisLysLeuIleProAsn 1366
QY 6148 ACAGAAATGCGAGCAGCAACAGAACCCGAGCAGCAGCAGCAACAAACGAAAA----- 6201
Db 1367 SerGluMetLeuHisLysThrLysGluMetLeuIleLeuTyrAsnGluGluLysValAsp 1386
QY 6201 ----- 6201
Db 1387 LeuLeuHisPheTyrValPheLeuProIleTyrIleLysAspIleTyrGluPheAsnIle 1406
QY 6202 ---GGAACATCAACAGCAACAAAGAAATCTACGTG---GGCACAAATGCTAAAGCC 6255
Db 1407 ValCysAspAsnSerLysThrMetTrpLysAsnGlnLeuGlyLysValIleTyrHis 1426
QY 6256 ATCTTTCGAATTAACAGATAGCAGGAGGTGAATAGAGGTGT----- 6300
Db 1427 IleThrValSerLysArgGluGlnLys-----ValLysGlyCysSerPheAspAsnGlu 1444
QY 6301 ---AATCCAAAAAGTATGAGCAATATCTTAATGGGCTGTATGATGATGATCT 6354
Db 1445 HisIleHisMetPheSerTyrAsnLysThrAsnValLysAsnCysIleIleAspAlaLys 1464
QY 6355 AAAGAAATGAATAATGCAATATGATGCTCTCTAGAGAAAAAATTA-----TGN--- 6405
Db 1465 ProLysAspLeuIleGlyPheValCysProSerGlyThrLeuLysLeuThrAsnCysPhe 1484
QY 6406 -----ATAAATATATACAAATTTTA----- 6426
Db 1485 LysAspAlaIleValHisThrAsnLeuThrAsnIleAsnGlyIleLeuTyrLeuLysAsn 1504
QY 6427 -----AATTAAT---GAACGTGAATAAGCGT 6450
Db 1505 AsnLeuAlaAsnPheThrTyrLysHisGlnPheAsnTyrMetIleIleProAlaLeuMet 1524
QY 6451 GACAATGATATAAAGAGCGCTTTATTAATGTGACAGCATATAGAACTCAATTTTGTGG 6510
Db 1525 AspAsnAspIleSer-----PheLysCysIleCysValAspLeuLys-----Lys 1539

QY 6511 TTAATAATATATGAAAAATCCTCGCAGCAAGAAATGATTCGAA----- 6555
Db 1540 LysIleuValHisasnGlyValAspThrGlyProValIleuAspAsnSerThrPheGlu 1559
QY 6556 -----AATGAAACAATTCAGATGATTAATTAAGAATA 6588
Db 1560 LeuAsnIleuLysPheAspAsnIleuValThrGlyThr-----AspGlnAsnIleuVal 1577
QY 6589 ATGTATATATACATATGCTGAT-----TATTAAGATATG 6621
Db 1578 Met-----ThrTyrIleAspLeuHisLeuSerHisLysArgAsnIleuValGluLeu 1595
QY 6622 TTTTTCGAACGTATATTTTGAATAAAAAATTTACTGTACAAATAGCTGTACA 6681
Db 1596 PheHis-----AspLeuGlyLysLysProAlaAspThrAlaAsnProGluSer 1613
QY 6682 ACCATT-----CTCATGAAATATATAGAAAAAACAGATTAATAAAAAAGAT 6729
Db 1614 IleIleGluSerLeuSerIleAsnGluSerAsnGluSerGlyProPheProThrGlyAsp 1633
QY 6730 GAAGAATTAACGTAATAATATTTTGGAGAAAAATTAATTTTGGAGAAAGATGATA 6789
Db 1634 ValAspIleGluHisLeu-----IleLeuGly----- 1643
QY 6790 TATGATTAATCATATCATCTCAGACAGCAAGAAAAAATTAAGATATATAC 6849
Db 1644 TyrAspThrTrpIleuSerLeuTyrAspGluGluLeuGluValIle----- 1659
QY 6850 CACTACATGACATGACCAACTG---ACGCTTCCTTCAGAGTTGTAAAAAGGCC 6906
Db 1660 ----TyrAsnAspIleGluSerLeuGluLeuLysAspIleGluGlnTyrValIleGluVal 1678
QY 6907 CAATTTTTCAGATGTTTCACAGAAATGGCAGAAATTTGTAAATAGCAGAAAGACAG 6966
Db 1679 AsnLeu-----LysAlaProLysLeu 1685
QY 6967 TTGTTAAATTTGGAGGGGCGCTGTAAGCATATAGTGTATATGATTAAGACGTAG 7026
Db 1686 MetIleSerAlaGlnIleHisAsnAsnAlaGlnValLysAspPheSerLysAsn----- 1703
QY 7027 ACACAGCAATGTCAGAGGCGTGTATACATATCAAAATTTTAAAGATGGCAAAACT 7086
Db 1704 -----AsnIleuIleValPro----- 1708
QY 7087 GAATATGAAGAACAAGAAAAAGTTCAAAAAGCATTAAGTGCACAAAAGTATAGGAT 7146
Db 1709 -----GluSerLeuLysLys----- 1714
QY 7147 TATCCTTCTACTGAAGACATAGACAGCAACATGCTCATGAATTTTAAACATG 7206
Db 1714 ----- 1714
QY 7207 AATTAAGAATATATGTCGAATAG-----GATGTCTCTGATCAAAAACCT 7257
Db 1715 -----GluIleuLeuGlyLysAsnProValAsnIleHisCysTyrAlaLeuLysPro 1732
QY 7258 TCTTCACACTA-----CAAAAAACAACACAAATCACAATCATCCGATGCT 7305
Db 1733 LeuAspThrLeuTyrValLysCysProThrSerLysAspAsnIleuValAlaLysVal 1752
QY 7306 AATGATATGCGCAGAAATCGCGATATATGTCGCAAGAAATTAACAAGTGTGATGCT 7365
Db 1753 Asn-----IleSerGlnAsnAspAsnGlyTyrGluLeuGln 1764
QY 7366 GAACCTTCAAAAAAGGATCATGATCATCAAAAAAATTAAGTGAACCTTAATACCT 7425
Db 1765 ValIleSerIleuIleGlyLysArgPheHisAsnPheGluThrLeuGluSerLysPro 1784
QY 7426 ATGATTTGTGTAGGAAGACGATATATTTATCTAAAGACGAAAAATATATGAT 7485
Db 1785 -----GlyAsnGlyAsp 1788
QY 7486 ATTACCTTGAAGAAAAATTTATA-----CCTATT-----GAGTCTACAAAGAA 7530

Db 1789 ValValAlaHisasnGlyValValAspThrGlyProValIleuAspAsnSerThrPheGlu 1808
QY 7531 AAGGAAGATTAATAATAGTGTGACTAATATATATTCCTGATCCTTAAGAAACCTTATGCA 7590
Db 1809 LysTyrPheLysAsnIle-----LysIleLys 1817
QY 7591 CCTGATTAATATATAGAGACA---ACAACCCCTTGCAAAATAGAGAGAAATCGTTTT 7647
Db 1818 ProAspLysPhePheGluLysValIleAsnGluTyrAspAspThrGluGluLys--- 1836
QY 7648 AAGGTACATATGAA-----TGGAAATGTTACAAA 7677
Db 1837 -----AspLeuGluSerIleLeuProGlyAlaIleValSerProMetLysValLeuLys 1854
QY 7678 AATTCAAAGTCTATACGAGAAAAAAGATATGTGTACTCCCAAGACAGAAACATATG 7737
Db 1855 LysLysAspProPheThrSerTyrAlaAlaPheValAlaProPro-----Ile 1870
QY 7738 TCGTAAAGATTTAGATGAATTAATAATGAA-----ACACTTAAGAGATAGT 7785
Db 1871 ValProLysAspLeu---HisPheLysValGluCysAsnAsnThrGluTyrLysAspGlu 1889
QY 7786 AATTATCTCTAAAAATGCTTCGTCGAACTGCAGCAAGAAATGACATATATATATA 7845
Db 1890 AsnGlnTyrIleSerGlyTyrAsnGlyIleIleHis-----IleAspIleSerAsn 1906
QY 7846 AACTCAACTCAGACAGAGGATGCGCAATGAAATCCATATGTGATTAATAATATAGT 7905
Db 1907 SerAsnArgLysIleAsnGlyCys-----Asp 1915
QY 7906 TTGCGTATGTCGGTGCATATAGTATAGACAGACATATGTCGAATGTCGTACTTA 7965
Db 1916 PheSerThrAsnAsnSerSerIleLeuThrSerSerValLysLeuValAsnGlyGluThr 1935
QY 7966 CTTCCCGTAGAATTAATAATTA----- 7986
Db 1936 LysAsnCysGluIleAsnIleAsnAsnAsnGluValPheGlyIleIleCysAspAsnGlu 1955
QY 7987 -----TATTAAGTTTGAATATATATATATGAAATGAGAAATTAATAAT 8031
Db 1956 ThrAsnLeuAspProGluLysCysPheHisGluIleTyrSerLys-----AspAsn 1972
QY 8032 AAAGGTAAATTAATCAACAGATGTA-----CAACGTTTCTGTCCTGG 8079
Db 1973 LysThrValLysLysPheArgGluValIleProAsnIleAspIlePhe-----SerLeu 1990
QY 8080 TGGGATGCTAATAGAAAGATTTTGAAGCA----- 8112
Db 1991 HisAsnSerAsnLysLysValAlaTyrAlaLysValProLeuAspTyrIleAsnLys 2010
QY 8113 -----ATGACGTGCAAAAGCACAGAGATGCAAAACTTTTAAAGAAAGAGA 8160
Db 2011 LeuLeuPheSerCysSerCysLysThrSerHisThrAsnThrIle----- 2025
QY 8161 ATGAGAGATTTGAACGCATATACATTAATACAGATAGTGTGCAATTAAGACAGATGCA 8220
Db 2026 -----GlyThrMetLysValThrLeu-----AsnLysAspGluLys 2037
QY 8221 CTTGTTGATATATATATACCTCAACGTTTCGATGATGACTGATGAATGTGATATAT 8280
Db 2038 GluGluIleAspPhe----- 2042
QY 8281 TGTAAAGCACTGATGAGAAATGGAATAATTAATAATCATGTGATCAC---TGTAAA 8337
Db 2043 -----LysThrAlaGlnGlyLysHisHisAsnValHisLeuCysAsn 2057
QY 8338 ACATCTACAGACGCAAGATGATTAATGATGAATAATAGTGTGAACGCTGATAA--- 8391
Db 2058 PhePheAspAsnProGluLeuThrPheAspAsnAsnLysIleValLeuCysLysIleAsp 2077
QY 8391 ----- 8391

Db 2078 AlaGluLeuPheSerGluValIleIleGluLeuProIlePheGlyThrLysAsnValGlu 2097
QY 8392 -----ACGAGATGTCAGAAATATATAAAATTTGTTCTTAATGGAATCTCATTCGAT 8445
Db 2098 GluGlyValGlnAsnGluGluIleLysLysPheSerLeuLysProSerLeuValPheAsp 2117
QY 8446 ATACAAATCAATTAATCAAAAGATTTGATACACCAATATATATCAAAAATCTCT-- 8502
Db 2118 AspAsnAsnAsnAspIleLysValIleGlyLysGluLysAsnGluValSerIleSerLeu 2137
QY 8503 -----ACTTATGATCATGTTCAAAATTTTGACAAAAGCTTGAAACATTTT-- 8547
Db 2138 AlaLeuLysGlyValTyrGlyAsn-----ArgIlePheThrPheAsp 2151
QY 8548 AAAAGTAAAGTCTGTGTGAGAGCTTTTCGAATATCTTCATGAAACAGTAAGTGTTC 8607
Db 2152 LysAsnLysLysLysGlyGluGlyIleSerPhePheIleProIleLysGlnAspThr 2171
QY 8608 AATATATAAATTT-----AATGAAAATGATGTTCTTCGAATATACGA----- 8649
2172 AspLeuLysPheIleIleAsnGluThrIleAspAsnSerAsnIleLysGlnArgGlyLeu 2191
Db 8650 ACATATGCTTTGAGAAACACCAAAAAAGTTATTAAGAACCTTGACGTGTACACTACT 8709
2192 IleTyrIlePheValArg----- 2197
QY 8710 TCATAAGATTCATTTGGATATTTGTTCTCCACCAATCAAAACAAAGATGATGAATTA 8769
Db 2198 ---LysAsnValSerGluAsn----- 2203
QY 8770 CAAACTTTTACCTTCTGCTGGAAGAAATGATATGATTAATATCTGATATATGGAACGA 8829
Db 2204 ---SerPheLysLeuLysAspPheThrThrGlySerThrSerLeu----- 2217
QY 8830 TACCTGTCTTATATGATGTCAGATGATTAACAAGGTGATGATTCCTCAACAGAGAGA 8889
2218 ---MetGluLeuAsnSerGlnValLysGluLys----- 2227
QY 8890 CATTATTTGATCAAGACCTTACATGCTGATATATATAGAAAGGTATTAAGAAATTTTA 8949
Db 2228 ---LysCysThrValLysIle-----LysLysGlyAspIlePheGlyLeu 2241
QY 8950 AAAAAAACTCTTACTTCTGCTTCACTCA-----GACAAATTTGATGCTCA 9000
Db 2242 LysCysProLysGlyPheAlaIlePheProGlnAlaCysPheSerAsnValLeuGlu 2261
QY 9001 AATATATAATCGAAGACATGTTGCTTGAAGCAATATATAGTATATGACATTAAT 9060
2262 TyrTyrLysSerAspTyrGluLysPheGluHisIleAsnTyrTyrIleHisLysAspLys 2281
9061 TCCGATATATTAAGAACTGAT-----ATGATGACACCTTCATTAATCTGAA-- 9108
2282 LysTyrAsnLeuLysProLysAspValIleGluLeuMetAspCysLysPheArgGluLeu 2301
QY 9109 -----AAAAATTAATAATTTGAAACATCAAT 9138
Db 2302 GlnAsnIleGluGlnTyrThrGlyLysSerAsnIleThrAspValLeuHisPheLysAsn 2321
QY 9139 GAAGCAACCGAATATCGTAACAATGTTGGGAAATATATAGACATATATGACACCT 9198
Db 2322 PheAsnLeuGlyAsnLeuProLeuAsnPheLysAsnHis----- 2334
QY 9199 ATGTATATGCGATTAATAATTTGCTACTCAAAAGTATAGATGAAGATGCTGTCAA 9258
2335 ---TyrSerThrAlaTyrAlaLysValPro----- 2343
QY 9259 TTACCAAGATGAGAAAGTAATATAGTTTCTGCTGTTATATGATGCGCAAGCA 9318
Db 2344 -----AspThrPheAsnSerIleIleAsnPheSerCysAsn 2355
QY 9319 GCAGTAAAGAAAGAAACATGATGATGATCATTAATAAACAATATGCTCGTTCAAC 9378
Db 2356 CysTyrAsnProGluLysHisValTyrGlyThrMetGlnValGluSerAspAsnArgAsn 2375
QY 9379 GAAGTAAATTTAGACGCTGAGATTAATTAAGACAACTGGATGTCAGATGATTAAG 9438
Db 2376 PheAspAsnIleLysLysAsnGlu-----AsnValIleLys 2387
QY 9439 AATATATATGCTTGAATATATGATTAATAAATATCAATGGAATATCAATTAATTAAT 9498
Db 2388 AsnPheLeuLeuProAsnIle-----GluLysTyr 2397
QY 9499 AACCAATTAAGAT-----CAATCTGAGTAAATATAGACAAATTAACATTCAGAGA 9552
Db 2398 AlaLeuLeuLeuAspArgGluArgGlnLysIleLysGlnGlnGlnGlnGlu 2417
QY 9553 AATGTCAGTATATTAATAATCAAAAGTTCTCAATGCGCTTGGAG----- 9600
Db 2418 GlnGlnGlnGlnIleLeuLysAspGlnAspArgLeuSerArgHisAspArgAsn 2437
QY 9601 -----TTAATGATATATAAGAAATGTTATACAGAACAAATAATAT 9642
Db 2438 LysAsnHisThrTyrIleLeuTyrAspSerAsnGlnHisIleCysAspTyrGluLysAsn 2457
QY 9643 GAAAT-----AATGAATTCAAAGAGTACTAAATAATATATATCT 9684
Db 2458 GluSerLeuIleSerThrLeuProAsnAspThrLysIleGlnLysSerIleCys--- 2476
QY 9685 GGTATATATTTGTTGAGATGAAACACAAATAATCATGTCATGATGAAATTAATA 9744
Db 2476 ----- 2476
QY 9745 GAAGAGACAAACAGTCTGCTTAAGCACTGATTTCTTTACA-----CCCAT 9795
Db 2477 -----LysIleAsnAlaLysAlaLeuAspValValThrIleLysCysProHis 2492
QY 9796 GTAGATCTTTCTATCAACGACCTTATTCACACATCGACATGACAAATATATGCT 9855
Db 2493 ThrLysAsnPhe-----ThrProLysAspTyrPhePro 2503
QY 9856 AAAAAATGATATATGAAAGTATCTCTGTTATGATGCGGCTAGGTTGATA 9915
Db 2504 AsnSerSerLeuIleThrAsnAspLysLysIleValIleThr----- 2517
QY 9916 GCGCTTATTCATGACAAAAAAATTCAAATCGTGTGAGCTGTGCGTATATGAT 9975
Db 2518 -----PheAspLysLysAsnPheValThrTyrIleAspProThrLys----- 2531
QY 9976 ATCCGCAAGAGAGTATGATGATGCTTGAATCCAAATAATAGTACATCATAT 10035
2532 -----LysThrPheSerLeuLysAspIleTyrIle----- 2541
QY 10036 AGAAGTGCATATTAAGCAAAACATATATATATGAGAGAGATAGTATGAGAT 10095
Db 2541 ----- 2541
QY 10096 GAAGTAAATATATGTTGGAATATCTTCTCGTATATTAATCTCATCCGAAGTAT 10155
Db 2542 -----GlnSerPheTyrGlyValSerLeuAspHisIleAsnGlnIleLysIleHis 2559
QY 10156 GAAGAAATGATATTAATGATATATATATATATGACAGGTAGCTCAATTAATAACATGATA 10215
Db 2560 GluGluThrPheAsp-----AspValHisLeu-----PheTyrProProHisAsnValLeuHis 2576
QY 10216 GAAGTACTAGTACAGCA----- 10233
Db 2577 AsnValValLeuAsnAsnHisIleValAsnLeuSerSerAlaLeuGluGlyValLeuPhe 2596
QY 10234 TCAAAAAGGATATACCAAGTATGATGATACACCAAGTAAATATACACACCTAGCAATAGA 10293
Db 2597 MetLysSerLysValThrGlyAspGluThrAlaThrLysLysAsn----- 2611
QY 10294 TTATATGATCATGAAGAAATGAAATGAAACATGATTTGTATCTCAATATTAACCAAT 10353
Db 2612 -----ThrThrLeuProThrAspGlyValSerSerIleLeu----- 2623

D	596	LysHisLeuLeuSerAlaLeuGluLeuLysLysGluLysGluLeuLysGluGlu	615
Q	4171	-----TTATTGGACAGCATATATGCAAAAGTCATGTGAGGAGTAACATAAAGAG	4224
D	616	LysThrLeuPhe-----SerCysTrpLysSerGluAsnGluLysLeuThr	631
Q	4225	CAATTAGATTCTCTTTTCAAATAATGGTGCACCAAAATATCTCCATTAAGGAAACACGCCCA	4288
D	632	GlnMetGluSerGluLysGluAsnLeuGlnSerLysIleAsnHisLeuGluLysLeu	651
Q	4285	GAATGTGACAGACAAACATCATGTAGATTTGGAGAGTCATGCTATGTCATAGATAA	4344
D	652	LysThrGlnGlnIleLysSerHisGluTyAsnGluAlaArgThrLeuGluMetLys	671
Q	4345	ATTGGGCGCAAAAAAGATGATTTTCCGAAACATACGGTTACACACSTCAAAATTAAT	4404
D	672	-----ArgGluAsnLeuSerValGluIleArgAsnLeuHisAsnVal---LeuLys	687
Q	4405	GACAAAGCACCACTTTGGAGGAATTTGCCAAACGACCCAGTTTATAGATGGCTAAC	4466
Q	4465	GAATGTGACGACGACATTTGCTATACACGACAAATATTTGAGAGTGCAGGAAAA	4524
D	708	GluPheSerLysPhe-----GlnLysHisGlnLysGluLysGluLysMet	721
Q	4525	TGTAACTCAATGACCAATGTAAGTGTGATACAGAAATGTAAATAGAAATGCAGAGATAC	4584
D	722	CysLeuLysThrSerGlnLeu-----ThrGlyLysValGlnLysPhe	735
Q	4585	GTT---AAATATATGAAAAAAGATGAGTATCCACAGATTAATATTACAGAT	4644
D	736	GluHisLysLeuGlnLeuLeuSerAsnGluIleMetAspLysAspArgCysTyrGlnAsp	755
Q	4641	-----	4644
D	756	LeuHisAlaGluTyrGluSerLeuAlaGlyAspLeuLeuLysSerLysAspAlaSerLeuVal	775
Q	4642	-----GAACGCGACAAAAAAGATTGATAGACA-----CAC	4674
D	776	ThrAsnGlnAspHisGlnArgSerLeuLeuAlaPheAspGlnGlnProAlaMetHisHis	795
Q	4675	ATTGGCTGAATGCTTACACAGCTATCTGACAGCAATGCAACAGATTAATTGAAACAGAA	4733
D	796	SerPheAlaAsnIleIleGlyGlnGlnGlySerMetProSerGluArgSerGluCysArg	815
Q	4735	TTTACTGCTAGTGTGTGTGATGAAGCGTGAAGCGCTCGTGGTGAACAAATATATACA	4794
D	816	LeuGluAlaAsp-----GlnSerProLysAsnSerAlaIleLeuGlnAsnArgValAsp	833
Q	4795	TTGTAGAA-----AAACAGCTTACTATGATCCGACCAAAATATGT	4836
D	834	SerLeuGlnPheSerLeuGlnSerGlnLysGlnMetAsnSerAspLeuGlnLysGlnCys	853
Q	4836	-----	4833
D	854	GluGluLeuValGlnIleLysGlyGluIleGluGluAsnLeuMetLysAlaGluGlnMet	873
Q	4837	-----GGTGGCAAAAATTTTGAATAATAGACAC	4866
D	874	HisGlnSerPheValAlaGluThrSerGlnArgIleSerLysLeuGlnGlnLysPheSer	893
Q	4867	AAATATATACATTAAT-----TCGATTAAGATTAAGTGCAAA	4902
D	894	AlaHisGlnAsnValAlaGluThrLeuSerAlaLeuGlnAsnLysGluLysGluLeu	913
Q	4903	GGATTATGTAAGAG---GCAAAACACAGCTGATTAAGTGCACAAACAAAGCTCTAAT	4955
D	914	GlnLeuLeuAsnAspLysValGluThrGlnGlnAlaGluIleGlnGlnLeuLysLysSer	933
Q	4960	AAATAC-----AATTAATCTGAAAAA-----TTGACGTAAGATGTGCTTTTCT	5004
D	934	AsnHisLeuLeuGlnAspSerLysGluLeuGlnLeuLeuSerGluThrLeuSerLeu	953

OY	5005	TCCTGCGACAGCTATATGTTTTCATGCAATGGACCAATTAATACAGATCCAGACTT	5064
Db	954	GlutylslysglutetserSerIlelleSerLeuasnLysargLileglu---GluLeu	972
OY	5065	AAAGATAAAATGGCTGCCAAAAGATTGATGGAAAGTCGGCGCAACGAGAGGTACAT	5124
Db	973	ThrcinglinsnglyThrIleuLysgluIleasnAlaserLeuasnGlnGluysmetasn	992
OY	5125	TTGGGCTAA-----TACTCAAAAGAAAAAGAAAAAGAGAAATA	5166
Db	993	LeuIleGlnLysSerGlnSerGlnSerPhealaasnTyrlleasnGluargGluysSerIleSer	1012
OY	5167	AAAAGTCGAGTCGCGACAAATATCTTATGAGAGTC-----CCGCTTCAGTCATG	5220
Db	1013	GluLeuSeraspGlnTyrlLysGlnGluLysIleuIleLeuLeuGlnIargCysgluGlnThr	1032
OY	5221	AAATATAGTTTTATGATTTTAAGAT-----	5247
Db	1033	GlyasnAlaTyrgluLysSerGlnLysTyrlLysAlaIleGlnGluLysasnSerLys	1052
OY	5248	-----ATAATTCAGGTATATGATTAATTTGGAAAGATGAAAAACAAAGACGAGCA	5298
Db	1053	LeuGluCysLeuLeuasnGluCysThrSerLecCysGlnasnTyrlLysasnGluLeuGlu	1072
OY	5299	AATTTGAGAAAAATATTTAAACAAAATGGAACATCAGTTGCCAAAGAAAGTATAGTACT	5358
Db	1073	GlnLeuLysGlnAlaPhealaLysGlnLHisGlnGluPhe-----	1085
OY	5359	ACAGGAATCCCGGTACTACTGCCGCCAAATTTTTTGGACGAAATATAG-----	5409
Db	1086	-----LeuThrLysLeuAlaPheAlaGlnGluIargGlnasnLeuMet	1100
OY	5410	-----GAATGTGTGGAACGAAATGATATGCGGCTACAAACGTCGTAGGATAT	5460
Db	1101	LeuGluLeuGlnThrValGlnGlnAlaLeu-----ArgSerGlnuMetThr	1115
OY	5461	CGAAATAGTGAATTAAGTCCACAGACTGATGAAGATCTAAAAAATTTGGTCTGTACTT	5520
Db	1116	AspasnGlnasnSerLysSerGlnAlaGlyLeuLysGlnGluIleuMetThrLeu	1135
OY	5521	TCACATGATGATATCTTATGGGGAAGAAAAATCCGATGAGAGTACGTCATCAGTCTT	5580
Db	1136	LysGlnGluGlnasnLysMetGlnLysGluValasnAsp-----	1148
OY	5581	CGATGTTTCCGCAATGGGTGAAGATTTTTCGAA-----CATTAAGAAAAGCAATTG	5634
Db	1149	-----LeuLeuGlnGlnLysasnGlnLeuMetLysValMetLysThrLysHisGluCys	1166
OY	5635	GAGCAATTCGTAAGGCGGTATATGATTAATCTTGGTGATTAATGAAGATTAAGAAG	5694
Db	1167	GlnasnLeuGlnSerGlnProIleargAsnSerValLysGluIargGlnSerGluIargAsn	1186
OY	5695	AAATGTCACAGTCGCTATACACAATTAATAAAATTTATTAAGATGAGATCGAACCACAGAT	5754
Db	1187	GlnCys-----AsnPheLysProIleMet	1194
OY	5755	GAAAAACAATCAAAAATATGTGTGAGAAATTAAGACAAAATATATATCCGACATCTGTG	5814
Db	1195	AspLeuGlnValLysGln-----Ile	1201
OY	5815	GCAAAAGATGACGAGCGCTCCGCAATATTTAAGCAAAACAATTAATAAAATTTGTGCA	5874
Db	1202	SerLeuaspSerTyrlasnAlaGlnLeuVal-----GlnLeuGlnAlaIleLeuIarg	1218
OY	5875	AATAAA-----ACTGAGATTTGGAATATATTAAGTATGATGAAA-----	5910
Db	1219	AsnLysGlnLeuLysLeuGlnGlnSerGlnLysGluLysGlnLysLeuGlnHisGluLeu	1238
OY	5911	-----GATGTGTCACACACAGCATTAATACGATGGATTAATAGTCAAAATATG	5955
Db	1239	GlnThrIleargGlnLysLeuGlnThrSerasnLeuGlnAsnMetClnSerCln-----	1256

QY 5956 CCCGATCATTTAGACATGACGACAAAAGATGAAAGAAAGTGTATTGTCAAGTCCCA 6015
Db 1257 -----GluIleSerGlyLeuLysAspCysGluIleAsp 1267
QY 6016 -----CGAGGTCCACCACTGTGA-----CGAAGGGAACACCGTCAACA 6054
Db 1268 AlAGluGluLysTyrIleSerGlyProHlsGluLeuSerThrSerIleAsnAspAla 1287
QY 6055 CGGATATCACTGATATCAAAAGCAGCGCATCGAAA-----AAGAGACGAAACAGCGCCG 6111
Db 1288 HlsLeuGlnCysSerLeuGlnThrThrMetAsnLysLeuAsnGluLeuGluLysIleCys 1307.
QY 6112 CCGACAAAACAGCCGAAAAGTGGAAATCTACACAGAAATG-----CGAGCA 6162
Db 1308 GluIleLeuGlnAlaGluLysTyrGlu-----LeuValThrGluLeuAsnAspSerArgSer 1326
QY 6163 CAA-----ACAGCAGCCGACGACGACACAAACGAAACGAAACGACATCAACAGCA 6216
Db 1327 GluCysIleThrAlaThrIleGlyMetAlaGluGluValGlyLysLeuLeuAsnGluVal 1346
QY 6217 ACAACAACAGAAATCTGACGTGGGACAAATGTAAGGCCATCTTTCGATAAACAGAT 6276
Db 1347 LysIleLeuAsnAspSerGlyLeuLeuHlsGlyGluLeuValGluLysIlePro----- 1365
QY 6277 AGCAGGGGTGAATAGAGGTGTATCCAAAACGATAGCAATATCTTAATGGGT 6336
Db 1366 -----GlyGlyGluPheGlyGluGlnProAsnGlu-----GlnHlsProValSerLeu 1381
QY 6337 TGTATTTGTAGTACTGTAAAGAAATGCAATGATGATGATGATGATGATGATGATGAT 6396
Db 1382 AlaProLeuAspGluSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1400
QY 6397 AAATATATGAT 6450
Db 1401 GlnMetHlsPheAlaGluLeuGlnGluLysPheLeuSerLeuGlnSerGluHlsIysIle 1420
QY 6451 GACATGATATATAAGAGGCTTTTATTAATGTGACGAAATGAAACGAAATTTTGTGG 6510
Db 1421 LeuHlsAsp-----GlnHlsCysGlnMetSerSerLysMetSerGlu 1434
QY 6511 TTTAAAT 6546
Db 1435 LeuGlnThrIleValAspSerLeuLysAlaGluAsnLeuValLeuSerThrAsnLeuArg 1454
QY 6546 ----- 6546
Db 1455 AsnPheGlnGlyAspLeuValLysGluMetGlnLeuGlnGlyLeuValPro 1474
QY 6547 GAATTCAGAAATGAAACATTCAGAT-----GAA 6576
Db 1475 SerLeuSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1494
QY 6577 TTTAAAGAT 6636
Db 1495 PheTyrAlaGlnAlaLeuLeuGlnGlnThrGlyAspMet-----SerLeu 1508
QY 6637 ATTCTATATGATTAATAAAT 6693
Db 1509 LeuSerAsnLeuGlnGlyAlaValSerAlaAsnGlnCysSerValAspGluValPheCys 1528
QY 6694 GAAATATAATAAAGAAACAGATTAATAAAGAAACAGATGA-----GAA 6735
Db 1529 SerSerLeuGlnThrIleValAspSerLeuLysAlaGluAsnLeuValLeuSerThrAsn 1548
QY 6736 TTACGTAAATATTTTGGGAGAAATATAAATTT-----ATTGGGAGAGATG 6786
Db 1549 LeuTrpAsnPheGlnGlyValAspLeuValLysGlnMetGlnLeuGlnGlyLeuValLeu 1568
QY 6787 AATATATGATTAAT 6843
Db 1569 ValProSerLeuSerSerSerSerSerSerSerSerSerSerSerSerSerSerSerSer 1588
QY 6844 AATTACAGTAC-----AATGACATGACCAAACTGACGCTTCC 6882

Db 1589 SerSerPheTyrArgAlaLeuLeuGlnGlnThrGlyAspMetSerLeuLeuSer-----Asn 1607
QY 6883 CTTCGAAAGATTTGTAAAAAGCCCAATTTTGTAGATGGTTCACAGATGGCAGAGAA 6942
Db 1608 LeuGlnGlyValAlaSerAlaAsnGln-----CysSerValAspGluVal 1622
QY 6943 TTTTGTATATAGAGAAAGACAGTGTGA----- 6972
Db 1623 PheCysSerSerLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1642
QY 6973 -----AAATGGAGCGGGCTGTAAAGATAT----- 6999
Db 1643 LysGlyValGluGlnLeuLeuSerLeuLysGluValAlaArgGlnSerLeuLysLeu 1662
QY 7000 -----GATGTATATGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7047
Db 1663 GluGluLysMetGlnSerGlnGlnIleMetLysAsnLysGluIleGlnGlnLeuGlnGln 1682
QY 7048 TGTATATATATCAAAAT-----TTTATTAAGATGGAAACGTAATAT 7092
Db 1683 LeuLeuSerSerGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1702
QY 7093 GAAAGACAAAGAGAAAGTCAAAAAG-----GATTAAGATGGCAAAAGTATAG 7143
Db 1703 GluGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1722
QY 7144 GATTTATCTTCATCAAGAA-----GACATAGAGAGAGCAACATGTGCTCATGAA 7194
Db 1723 GluLysLysGlnThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 1742
QY 7195 TATTAACATGAAATTAAGAT 7227
Db 1743 GlyLeuAspLeuSerSerArgSerLeuGlnGlyLeuAspThrGlnAspAlaIleGlnGly 1762
QY 7228 ---AATAAGATTTGTTCTGTATGCAAAAACCTTTTACAACTACCAAAACAAACAA 7284
Db 1763 ArgAsnGlnSerCysAspIleSerLysGlnIleThrSerGlu-----ThrThrGlu 1779
QY 7285 CAATCACATATCCGATGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7344
Db 1780 ArgThrProLysHlsAspValHlsGlnIle-----CysAspLysAspAlaGlnGln 1796
QY 7345 TTTAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7404
Db 1797 AspLeuAsnLeuAspIleGlnLysIleThrGlnThrGlyAlaValLysProThrGlyGlu 1816
QY 7405 ATTACTGAACCTAAATACCT----- 7425
Db 1817 CysSerGlyGlnGlnSerProAspThrAsnThrGluProProGlyGluAspLysThrGln 1836
QY 7426 -----ATGAATTTGTGTAGAAAGACAGCATAT----- 7452
Db 1837 GlySerSerGlnCysIleSerGlnLeuSerPheSerGlyProAsnAlaLeuValProMet 1856
QY 7453 ---TATTTATGTAAGAAGACGAA---AATATATGATATATACCTTGAG----- 7497
Db 1857 AspPheLeuGlnAsnGlnGlnAspIleHlsAsnLeuGlnLeuAsnGlnValLysGlnThrSer 1876
QY 7498 ---GAAATATTTATACCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 7554
Db 1877 AsnGlnAsnLeuThrGlnLeuHlsValIleGlnAspArgAspArgLysValGlnSerLe 1896
QY 7555 AATATATATCTTCGATGCTATGAAACCTTATGACCTGTAATATATATATATATATATATATAT 7614
Db 1896 uLeuAsnGlnMetLysGlnLeuAspSerLysLeuHlsLeuGln-----GluValGln 1913
QY 7615 AACCTTTGTGAATATAGAGAAAGAAATCGTTTAAAGTATGATGATGATGATGATGATGATGAT 7674
Db 1913 nLeuMetThrLysIleGlnAlaCysIleGlnLeuGlnLysIleVal---GlyGlnLeuLys 1932
QY 7675 AAAATTCAAAGTCTATCAAGAGAGAAAGAAAGATATGTACTCCAAAGAGAAACAT 7734

Db 1932 slysgluasnseraspleuSergluylsleuGluTyrPheSerCysAspHisGlnIle 1952
QY 7735 ATGCTGTTAAGCAATTTACATGCAAAATTAAGAACTTAAGATGTAATATATC 7794
Db 1952 - - - - -LeuGln- - - - -ArgValGluThrSergluLeuSerSerAsp 1965
QY 7795 CTAATAATGTTGCTGACACTGACGAAATGAAGAAATGACATATATAAAAACTTCAC 7854
Db 1966 LeuGluMetHisAlaAspLysSerSerArgGluAspIle- - - - - 1978
QY 7855 TCAGAGACGGGTGCGCAGATGCAATATGATATGATATGTAATATGTTGCTGAT 7914
Db 1979 GlyAspAsnValAlaIleValasn - - - - -AspSerTriPylSerGluArgPheLeuAsp 1995
QY 7915 CTGGGTGACATAGTTAGAGAACAGATATGTAAGATTTGGTGGTACTTACCTCCGTA 7974
Db 1996 ValGluAsn- - - - -GluLeuSerArgIleArgSergluLysAlaSerIle 2010
QY 7975 GAATATAATTTATATAGTTTGTGAAATACATATATGGAATGGAATATAAAAAATAA 8034
Db 2011 Glu- - - - -HisGluAlaLeuTyrLeuGlu- - - - - 2018
QY 8035 GGTAGAAATAAATACACATGATACAAAGTTTCGTTGCTGCTGGATGCTAATAGA 8094
Db 2019 - - - - -AlaAspLeuGluValValGlnThr- - - - -GluLysLeuCysLeuGluLysAspAsn 2035
QY 8095 AAGATATTTGGAACCAATGACGTCGCAAGACAGAA- - - - -GATGCA 8139
Db 2036 GluAsnLysGlnIleValIleValCysLeuGluGluIleuSerValIleThrSerglu 2055
QY 8140 AAACCTTTTGAAGAGAAATGATGATTTGAACGATATACATTAATACAAAGTAAG 8199
Db 2056 ArgAsnGlnLeuArgGluLeuAspThrMetSerLysLysThrAlaLeuAspGln 2075
QY 8200 TGTGACATTAAGACACATCCACTGTGATGATATATACCTCAACGGTTTCGATGATG 8259
Db 2076 LeuSerGluLys- - - - - 2079
QY 8260 ACTGAATGCTCGAATATTTATGTTAAAGCACTGATGAGAAATTTAAATAA 8319
Db 2080 - - - - -MetLysGluLysThrGlnIleuGlnIleuSerHisGlnSer 2092
QY 8320 TCATGTCATGCTGT- - - - -MetLysGluLysThrGlnIleuGlnIleuSer 2092
Db 2093 GluCysLeuHisCysIleGlnValAlaGluAlaGluValLysGluLysThrGluLeu 2112
QY 8350 TGCAGCAATGATATGATGAAATTAAGTGAACAGTGTAAAGCAGATGTCAAGATAT 8409
Db 2113 GlnThrLeuSerSerAspValSergluLeuLysAspLysThrHisLeuGlnIleLys 2132
QY 8410 AAAAATTTTGTCTTAATGAAGAAATCTCTATTCGATATA- - - - - 8448
Db 2133 LeuGlnSerLeuGlnLysAspSerglnAlaLeuSerLeuThrLysCysGluLeuGluAsn 2152
QY 8449 - - - - -CAATCAATTAATACAAAGAAATGTAT- - - - -GACACCAATATAT 8490-
Db 2153 GlnIleAlaGlnLeuAsnLysGluLysGluLeuValLysGluSerGluSerLeuGln 2172
QY 8491 ACAAAATCTCTACTTATGATCATGTTCAAAATTTGTACAAAGTGTAAACCTTTTAA 8550
Db 2173 AlaArgLeuSerGluSerAspLys- - - - -GluLysLeuAsnValSerLys 2187
QY 8551 AGT- - - - -GAATGTTCTGTGACAGCTTTTGAATATCTTCATCAACAAAGTAAAGTGG 8607
Db 2188 AlaLeuGlnAlaIleValValGluLysGluPheAlaLeuArgLeuSerSerThrGln 2207
QY 8608 AATTATAATTTATGAATAATGATGTTCTTCAATATGCAACATATGTTTCGAAGA 8667
Db 2208 GluGluValHisGlnLeuArgGlyIleGlnLysLeuArgValAlaArgIleGluAlaAsp 2227
QY 8668 ACACCAAAAGTTTAAAGACCTTGCAAGTGTACACTACCTTCTAAGATTCATTTGAT 8727
Db 2228 GluLysLysGlnLeuHisIleAla- - - - - 2235

QY 8728 AATTGCTTACGATCAAAACAAAGATGATTAAGCAATTAACAACCTTTACCTTTCG 8787
Db 2236 - - - - -GluLysLeuLysGluArgGluArgGluAsnAsp- - - - - 2246
QY 8788 TCAGAAATGATATATATATATATCTTGATTAATGGAAGCAATCTTGTCTTAATAGT 8847
Db 2247 SerLeuLysAspLysValGluAsnLeuGluArg- - - - -GluLeuGlnMet 2261
QY 8848 TCAGATGATTAACAAGGTGATTTGATTTCTCCAAAGAAAGACATTTATGTAACAGCT 8907
Db 2262 SerGluGluAsnGlnIleuValIle- - - - - 2270
QY 8908 ATCAGTCATATATATTAAGAAAGTGAATAAGAAATTTTAAAAAACTT- - - - - 8961
Db 2271 LeuAspAlaGluAsn- - - - -SerLysAlaGluValGluThrLeuLysThrGlnIleGluGln 2289
QY 8962 CTTCCTTCGCTTTCAGTCAAGACAAATGTTAGTCGCAAAATATTAATCCGAAGAAAG 9021
Db 2290 MetAlaArgSerLeuLysValPheGluLeuAspLeuValThrLeuArgSergluLysGlu 2309
QY 9022 TTGCTGCTTGAGGCAATGAATATAGTTATGCAATTTATCCGATTAATTAAGAGACT 9081
Db 2310 AsnLeuThrLysGlnIleGlnIleuLysGlnGlnLeuSerGluLeuAspLysLeu 2329
QY 9082 GATATGATGACACTTCATTTATCTGAATAATTA- - - - -AAATATTTGAAGA 9132
Db 2330 SerSerPheLysSerLeuLeuGlnGluLysGlnGlnAlaGluIleGlnIleLysGlnGlu 2349
QY 9133 TCAAATGAAGCAACGAAATGCTAAACATGTTGGGAAATATATAGCTGCAATAG 9192
Db 2350 SerLysThrAlaValIle- - - - -MetLeuGlnIleuGlnIleuLysGluLeuAsn 2365
QY 9193 CACGCTATG- - - - -TTATGTGATATAA- - - - -ATTGCTACTTCAAAAGTATACATTA 9240
Db 2366 GluAlaValAlaIleLeuCysGlnIleGlnIleMetLysAlaThrGluGlnIleuSer 2385
QY 9241 GATGAAGATGTTGTCATATTAACCAAGATGAAGAACTATATAGTTT- - - - - 9288
Db 2386 Asp- - - - -PropiolIleGlnIleuGlnIleuHisGlnIleuArgAsnSerIle 2399
QY 9289 - - - - -CTTCGTTGTTAATTTGA- - - - -TGGCAAGCAACGATCT- - - - - 9324
Db 2400 GluLysLeuArgAlaArgLeuGlnAlaAspGluLysGlnIleuLysCysValLeuGln 2419
QY 9325 - - - - -AAGCAAAAGCAATGATGATCTATTA- - - - - 9357
Db 2420 LeuLysGlnSerGlnHisAlaAspLeuLeuLysGlyArgValGlnAsnLeuGluArg 2439
QY 9358 - - - - -ACAAATGTCCTGTTCAACGAAGATAT- - - - -TTGACGCTCA- - - - - 9399
Db 2440 GluLeuGlnIleAlaArgThrAsnGlnIleuHisAlaLeuGlnIleGluAsnSerLys 2459
QY 9400 - - - - -GAATTTATGACAA- - - - -CTGAGTGCAGAAATGATTAATAAATAT 9444
Db 2460 GlyGluValGlnThrLeuLysAlaLysIleGlnGluMetThrGlnSerLeuArgGly- - - - - 2478
QY 9445 ATTAGCTGATATATGTTG- - - - -ATAAAAATTAACAATGAAATCTA- - - - - 9486
Db 2479 LeuGlnLeuAspValIleThrIleArgSergluLysPheuThrAsnGlnLeuGln 2498
QY 9487 - - - - -AATATAATTAAGCAATTAAGATTAAGATCAATCTTCAGGTAATATAGAC 9534
Db 2499 LysGluGlnGluArgIleSergluLeuGlnIleLysAsnSerPheGluAsnIleLeu 2518
QY 9535 AATTAACATCTCAGAGAAATGTTCAAGTCAATATTAATAAAGATTCATCAATGCGCT 9594
Db 2519 GlnGluLysGlnIleuLysValGlnMetLysGlnLysSerThrAlaMetGluMet 2538
QY 9595 TTG- - - - -GAGTTAATGATATATTAAGAAATGATTAACAGCAACAAATAAT- - - - - 9642
Db 2539 LeuGlnThrGlnLeuLysGlnLeuAsnGluArgValAlaAlaLeuHisAsnAspGlnGlu 2558

OY	9643		-----GAAATATTGAATTCAGCAAGAATA-----CTAAAAAAA	9675
Dd	2559	Alaclyslalylalslgluglnasnlseusercinvalilucysleugluleglnls	:	2578
OY	9676	-----TTATATCCCGTGTATATATTTGGTAAGATGACAAACAAMATCATGTGA---		9726
Dd	2579	AlaInLeuLeuIngnIingnIeU---AspClunlalylasaAenAnTYrlleVal		2594
OY	9727	-----CTAGATGGAATATATAAGACAGAACAG	:	9753
Dd	2595	LeugInSerSeRvaInasnIglyleulleIgInluValGlu	:	2607
<hr/>				
RESULT 6				
PVDB_PLAKN	ID	PVDB_PLAKN	STANDARD;	PRT; 1153 AA.
AC	P50493;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Duffy receptor, beta form precursor (Erythrocyte binding protein).			
PLASMODIUM KNOWLEDGE:	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.			
NCBL_TaxId=5850;	[1]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=92357776; Pubmed=1496004;			
RA	Adams J.H., Slim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;			
RT	"A family of erythrocyte binding proteins of malaria parasites.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992)			
CC	-!- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP DETERMINANT.			
CC	-!- SUBCELLULAR LOCATION: Type I membrane protein.			
CC	-!- SIMILARITY: HIGH TO P.VIVAX DUFFY RECEPTOR.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isdb.ch/announce/ or send an email to license@isdb-sib.ch).			
DR	EMBL; M90694; AAA29603.1; -			
KW	Malaria; Receptor; Glycoprotein; Signal; Transmembrane; Multigene family.			
FT	DOMAIN	1	21	POTENTIAL.
FT	SIGNAL	1	21	POTENTIAL.
FT	CHAIN	22	1153	DUFFY RECEPTOR, BETA FORM, EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	22	1085	POTENTIAL.
FT	TRANSMEM	1086	1106	CYTOSOLSMIC (POTENTIAL).
FT	DOWAIN	1107	1153	POTENTIAL.
FT	CARBOHYD	134	134	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	179	179	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	202	202	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	252	252	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	348	348	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	430	430	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	467	467	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	476	476	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	576	576	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	626	626	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	722	722	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	847	847	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	856	856	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	900	900	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	910	910	N-LINKED (GLCNAC . .) (POTENTIAL).
FT	CARBOHYD	935	935	N-LINKED (GLCNAC . .) (POTENTIAL).
SO	SEQUENCE	1153 AA;	130471 MR;	6497BD16CE7BEC01 CRC64;

Alignment Scores:	
Pred. No.:	1,79e-09
Score:	349.00
Percent Similarity:	34.4%
Best Local Similarity:	19.47%
Query Match:	1.79%
Length:	115
Matches:	259
Conservative:	199
Mismatch:	509
Indels:	363

DB: 1 Gaps: 59

US-10-087-013-1 (1-10628) x PVDE_PLAKN (1-1153)

QY 3010 GAAGCTAATATGAGACCAAGTATGGAAGCCATGAAA-----TGTGAT 305
Db 20 LysAlaAsnAsnVal---LeupheglIurghetleuGlylleLeuleuLeuGlySglu 38
QY 3052 ATATAATATTTGAGAGATTAATCCGGACACCAATCAACACAACTACTATTTCGGATAT 311
Db 39 AsnGluTyrValLysAsnGluAsnGlyTyrLysLeuAlaThrGlyHisHis-----Tyr 56
QY 3112 AGTATCATATACACCATTTGATGATTAATATCCACAAAAATTAAGATGATGACCGAATG 317
Db 57 MetAspAsnAspGlnIleGlu-----ArgTyrLeu-----GlnGly 68
QY 3172 GCAGATGTGCTACTGCAGAGGTGCAGAAAAGAGATATGATTAAGTGAAGCAGAACTGAAG 323
Db 69 ThrAspArgSerArgArgValLysIleGlnGluAsnValLysTyrLysTyrAsnValGlu 88
QY 3232 GAGTGAAGAGATAAGATAATGGTCAAGCGTGTACGAAGAGAGTGGTACAGTTGTAG 329
Db 89 GluLeuAsn-----ThrsTyrGluGlnThrLysGlyLys 100
QY 3292 AAGTCACACGAAGCTTGTAAATGAATATATAGATTAATAGATTAATGGAAGACAAATGG 335
Db 101 ArgTlleAsnArgIleLeuLeuGlySgluSerThr-----TyrGluAlaGln 114
QY 3352 AATATATATATCGATGAATATCAAGAAATTTACATGCACACACAAATGCTGTGTAGTAA 341
Db 115 AsnValAlaAspAsnAsnTyrIleAspSpLysAlaAsnGlyLysThrAspAsn 134
QY 3412 ACTGCTATTGAAGCTTCACAGACTGCCAAAAATCATATAGACAGAGATGTATATGAATTT 347
Db 135 -----LysThrAsnLysGlyGluGlyLysAlaArgGlnMetValMetLeu 148
QY 3472 TTGTGGAATTAATACCAAAATATGGTGGCAAAAGTAATATTAAGTGTACTAGTGA 353
Db 149 AspTyrAspIle-----SerGlySerGlyHisProAspIlylleIleAspAsn 164
QY 3532 ACTGCTGTCAATTTGGTACTAACACACACGATGAATAAATGTGGACATATCTCCATATCA 359
Db 165 ValValAlaGluLeuGlyThr-----GluAspGlu 173
QY 3592 GGAATTTTGAATGATTTGTCAGTCCAGTCCAAAAATGAATTTTGTGATGAAAAAAGTATGTA 365
Db 174 GlyAsnPheLeuGluAsnSerSer----- 181
QY 3652 GATTAACGAAAAATATGCTTATAGATTAACACAGACGACCAATGATGTGCGTGTGT 371
Db 181 ----- 181
QY 3712 AAAAGTGCATGAAACCGACAGAGGTACAGATTAACGAAAAAAGCGGAGCAAAAG 377
Db 182 LysGlyGlyAspHisProTyrArgMet-----AsnArgLys 193
QY 3772 GATACGCGAATGATAA---ACAGTGAATGATATCTTAAAGAAAGAGATGGAAGAAACA 382
Db 194 GluArgMetLysSerGlyValIleAsnGlnThrPheLeuGln-----LysAsnVal 210
QY 3829 GTAGAAAGATTTGTCATCCAAAAAAGATTAATGATATCCGATTTGGCAATGCGCGAAT 388
Db 211 MetArgArgCysAsn---AsnLysArgLysArgGlyThrArgAspTyrAspSerProThr 229
QY 3889 ATTAATTTTAATGGAAAGACCTCGTGTGTATGCCCCCTAGAAAGCAAAAGTATTCGCTA 394
Db 230 -----LysLysAspValCysIleProAspArgArgTyrGlnLeuCysMet 244
QY 3949 CATTTCCTGGCAAT-----GATTAATGATTAATAAAA 398
Db 245 LysGluLeuThrAsnLeuValAsnAsnThrLysThrHisSerHisAsnAspIleThrPhe 264
QY 3982 TTACATACACAACTTAATTTAAAAAGACCTTTCATCCAAATGTGCAGCAGCAAAACATTC 404

Db 265 Leu-----LysLeuasnLeuLysGluLysLeuThrTyrAspAlaIleValGluLys 282
QY 4042 TTCATGCGATATATATAAAAGTAAGATCGTGAAGAAATGACCTGCATTAAGATTA 4101
Db 283 LeuLeuLeuLysLysTyrAsnAsnValTyrSerGlu-----AspLeuLysAspIle 300
QY 4102 AAGAAGGCAAAATTCCTCCCGCATTTTGAGATTCACATGTTGACATTTGAGATTA 4161
Db 301 Lys-----TyrSerLeuGluAspPhe 307
QY 4162 AGGATTTTATTTGAGACAGATATATCAAAAGTCATGTCGAGGAGTAACATTA 4221
Db 308 GlyAspIleIleMetGlyThrAspMet--GluGlyIleGlyTyrSerGluValGlu 326
QY 4222 GACCAATAGATTCCTCTTTCAAAAATGTCGACCAAAAATCTCCTATGCAAAAACGCG 4281
Db 327 AsnAsnLeuArgThrValPheGlyThrGlyThrLysThrGluLeuAsp-----Arg 343
QY 4282 CAAGAATGTGACAGACATAGTACAGATATGGAATGGAACCTATGCTATGTCATGTA 4341
Db 344 LysLysTyrPheAsnGluSerLysLysTyrIleTyrGluIleThrIleLeuSerValLys 363
QY 4342 AAAATTGGGCAAAAAGATGATTTTACGAAACCTAGCGTTACAC-----AACGTC 4395
Db 364 Lys-----LysLeuAsnGlyTyrSerAlaTyrPheAsnCys 374
QY 4396 AAATTTAGTCAAAAAGCACACCTTTGAGGAATTTGCCAAACGACCCGATTTTTCAT 4455
Db 375 Lys--GluAspValGluIleAsnValGlu-----ProGluIleTyrArg 388
QY 4456 TGGCTAACGATGTGACAGACGACTTGTCTATACAGCAAAAATTTGGAAGATGTG 4515
Db 389 TrpIleArgGluTyrPheGlyMetAspTyrMetSerGluLeuProLysGluIle 408
QY 4516 CAGGAAATGTAAG-----TCAATGACCAATGTAAGTGTATACGAA 4560
Db 409 LysGluLysCysAspArgLysLeuTyrTyrThrAsnLeuArgIleCysThrMetSerPro 428
QY 4561 TGTATTAAGAAATCGAGACACTAGTAAATATATG-----AAAAAAGAGATGAT 4617
Db 429 CysAsnAspSerCysLysLeuTyrAspGluTrpIleThrArgLysLysGluTrpAsp 448
QY 4618 CCAACAGATATATTCACAGATGACGACGACAAAAAAGATTCATACACACAT 4677
Db 449 ValLeuSerThrLysPheSerSerValLysGlyGluIleIleGluThrGluAsnIle 468
QY 4678 GGTGTATGTTACAGACTTACTGGAAGCAATGCAACAGATTAAGTGAACGAAATTT 4737
Db 469 -----ThrThrAlaTyrAspIleLeuLysGluIleLeu 479
QY 4738 ACTGCTAGTTGTGATGAAGCCTGGAAGTCCCTGTGCTGACAAAGAAATATACATTTG 4797
Db 480 Asn-----GlyPheAsnGluValMetPheGluAsnGluIleAsn 492
QY 4798 TTGAAAAACAGGCTTACTATGATGCCGACAAACATTTGCTGGTCACAAATTTATTTGA 4857
Db 493 LysArgAspAsnValTyrIleAspIle-----CysLeuLysAlaIleAspGluPro 509
QY 4858 AATGACGACAAATATACTACATTTGAGTAAGATTAAGTGAAGATTAAGTAAGAG 4917
Db 510 AsnLysAsnThrGluIleGluIleLysLysLeuLysSerAlaProLysLeuGluThrGlu 529
QY 4918 GCAAAACAGCTCTATTAAAGTGGCAAAACAAAGTCTTATTAATCAATTAATTTGAA 4977
Db 530 ArgSerIleSerThrIleGlu-----ProMetSerSerSerGlyAlaGlu 544
QY 4978 GAATGCTAGATGATGCTTTTCTTCCTTCGTCGACTACGATATATTTTCATGACTTG 5037
Db 545 LysValGluIleLysAspLeu-----Ala 551
QY 5038 GATGCAATTAATACAGATCCAGAAAGTTAAAGATGAAATGGTTGCAAAAAGATGATG 5097

Db 552 HisGlyAsnIleAsnAspAlaIleTyrLysSerThrThrAsp----- 565
QY 5098 GAAGTGGCCGACACGAGAGGTAATTTGGTCATATCTACAAAGAAAAA 5151
Db 566 GluAlaIleLysGlyAspGlyGluAsnGlyLysAsnGluThrValAlaGluSerAsnIleLys 585
QY 5152 -----GAAAAAGAAAAATTAAGACGTGGATCGACGACAAATAT 5190
Db 586 GlyThrAspAsnIleGluAsnGluAlaIleAlaLysAsnValAspThrTyrLysPheValThr 605
QY 5190 ----- 5190
Db 606 GluArgSerAlaAspThrArgIleValAlaThrAspIleThrGlyThrGluLysLeu 625
QY 5191 -----TCTATAGAGTCCCGCTGTAGTGCATGAATATATGTTTATGAT 5238
Db 626 AsnThrSerTyrSerGlySerSerGluIleThrValLysLysAsnIleProGlyAspGly 645
QY 5239 -----TTAAGATATATTTAGTATTTGATTAATTTGAGATGAAAAAAGACCGAG 5295
Db 646 IleValLysAspValSerAlaIleValGluAsnSerGluAsnProLeuGluThrLysHis 665
QY 5296 GAAATTTGAGAAATATTTAACAACAAATGCAATCAGTTGCAAGAGAGT 5349
Db 666 LysIlePheGluProSerLysAspAsnSerAspAsnSerGluAsnSerGlySerMetGlu 685
QY 5350 ---GATAGTACTACAGAAATCC-----GATAGTACTGCGCGAAA 5388
Db 686 PheLysAlaThrSerSerAsnProIleThrGluAlaValGluSerSerAlaGluGly 705
QY 5389 TTTTTCGACAGCAAAATTAAGATGTGTGTGACACCAATGATATGCGGCTCAACAGT 5448
Db 706 GluValGluIleLysPheSerAlaHisArgSerValAsnThr----- 718
QY 5449 GGTAGGATGATGCAATTAAGTGAATGCAAGAAAGATGATGATGAT 5499
Db 719 GlyArgAspAsn---SerThrIleSerAlaIleThrSerAspAspGlyLeuSerSerGly 737
QY 5500 ---AAAAATGTGTTCTGTACTTACATGATGATGATATTCATGCGGAAAAATCCGAT 5556
Db 738 AspLysArgValGluSerIleLeuThrSerIleGluAsnAlaAspAspGlyLysProVal 757
QY 5557 GAAGTACTGCTGATTCAGTTCTTCA-----TGGTTCCGGAATGGGTGAAGAT 5607
Db 758 GluLysSerLeuThrPheAsnLeuAsnAspProSerValGluAlaGlyGlyLysSer 777
QY 5608 TTTTGCAAACATTAAGAAAGAAATTCGAGAAATGTGACGCGGTATGATTAATACT 5667
Db 778 HisIleLysThrGluGluAsnGlu----- 785
QY 5668 TGTGCTGATATGAAGATTAAGAAAGAAAGATGACAGATCGTGCACAAATTAATAAAA 5727
Db 786 ---GlySerGluAlaGluIleAspGlyLysAsnValAspIleAlaGluIleArgThrAla 804
QY 5728 TTTATTTAGTGAAGTGAACACACAGTAT-----GAAAAAACAATC 5766
Db 805 ThrIleThrGluValGluProGluArgProAspLeuSerAspThrAspAsnGlyAsnVal 824
QY 5767 AAAAAATATGCTGAGATTAAGACAAATATATTTCCAGCATCCTGCGCAAAAAGATGCA 5826
Db 825 ProArgSerGly---AsnLysGlu-----AsnGluGlyAlaThrAlaLeuSerGly 840
QY 5827 GAGAGCGCTCGCAATATTTAGACAAACAAATTAATAAAATTT 5868
Db 841 AlaGluSerLeuGluSerAsnGluSerValHisLysThrIleAspAsnThrThrHisGly 860
QY 5869 TGTGAATAATAAGTGAAGATTTGGAATATTAAGTGTGAAGATGTGTCACACAGCA 5928
Db 861 LeuGluAsnLysAsnGlyLysAsnGlu-----LysAspPheGluLysHisAsp 876
QY 5929 TTAACGTAGTAAATAGTCAAAATATATCCCGCATCATTAACGTGAACCAAAAGAGT 5988
Db 877 PheMetAsnAsnAspMetLeuAsnAspGluThrSer---SerAspGluThrSerSerAsp 895


```

QY 5989 GAAGGAAGTAAATTCGAAAGTCCACGAGTCCACCGCTGACGAAGGGAACACCG 6048
DB 896 GlnThrSerSerAsn--GlnThrSerSerAspGlnThrSerSerAsnGlnThrSerSera 915
QY 6049 TCACCCAGGGATACATGATTCAAAGCGGCGCATCGAAAAAGAACGAAACACGCG 6108
DB 915 spGlnThrSerSerAspGlnThrSerSerAsp--GlnThrSerSerAspGlnThrS 933
QY 6109 CCGGCTTCAAAACAGCC-----GAAAGAAAGTGAANAATCTAACA 6147
DB 933 eSerSerAsnGlnThrSerSerAspGlnThrSerAspThrGlnGlnGlnHisArgAspAsnV 953
QY 6148 ACAGAAAT-----GCAGACA 6162
DB 953 aAlaArgAsnProGlnThrLeuLeuSerSerGlnAspMetSerLysGlnAspPheMetArgAsn 973
QY 6163 CAACACGACACCGCAGC-----AGCAGACACACACACAGAAAGCAACAGACACACA 6213
DB 973 eAsnSerSerAsnGlnThrSerSerHisAsnAsnLeuAsnAsnArgLysLeuAsnGlnLeuAsn 993
QY 6214 GCACACACACACAGATTCAGCTGGCGCAATGTAAGGCCATCTCTTCG----- 6264
DB 993 IntYrGlnHisArg-AspValLysAlaThrArgGlnLysLysLeuLeuMetSerGlnVal 1012
QY 6265 AATAAACACATATGACAGGGGTGGAATAGAGGTTGTAATCCAAACGTAATGACACATAT 6324
DB 1013 AsnLysCysAsnAsnArgLysSerLeuLysTyrCysAsn----- 1025
QY 6325 CCTAAATGGGGTGTATTTAGCTAAGTCTAAGTAAAGAAATGAA-----AATGCATATGT 6378
DB 1026 -----ThrLeuGlnAspArgMetLeuSerSerThrCys 1036
QY 6379 ATGCTCTCAGAGGAAAAAATTA-----TGATTAATAATATATA 6417
DB 1037 SerArgGlnThrSerLysAsnLeuCysCysSerLysSerAspPheCysLeuAsnTyrPhe 1056
QY 6418 CAATATTTAATTAATGAACCTGAATAAGCGTGAACATGATTAAGAGCGTTTAT 6477
DB 1057 GlnLeuTyrProThrGlnThrLysPheThrCysMetLysGlnPheGlnHisSerSerTyr 1076
QY 6478 AATATG-----GCAGCAATAGAA 6495
DB 1077 GlnCysPheThrLysGlnSerSerThrGlyLeuGlyLeuValTyrPheAlaThrGlyGly 1096
QY 6496 ACTCAATTTTGGTAAATATATATTAATGAATCCCGACGCAAAATGAATGCA 6555
DB 1097 AlaPheLeuLeuLeuLeuLeuLeuPheValSerLysAsnValAlaSerAsnAspTyrGln 1116
DB 6556 AATGACAAATTCAGATGATTT 6579
DB 1117 GlnGlnAlaThrPheAspGlnPhe 1124

```

```

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH. TO P.KNOWLEDGE DUFFY RECEPTORS.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M61095; AA63423.1; -
KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane.
FT SIGNAL 1 1070
FT CHAIN 21 1070
FT DOMAIN 21 1007
FT TRANSMEM 1008 1025
FT DOMAIN 1026 1070
FT CARBOHYD 183 183
FT CARBOHYD 255 255
FT CARBOHYD 351 351
FT CARBOHYD 420 420
FT CARBOHYD 715 715
FT CARBOHYD 787 787
FT CARBOHYD 825 825
FT CARBOHYD 903 903
FT CARBOHYD 938 938
SQ SEQUENCE 1070 AA; 119683 MW; CB051D13E294603 CRC64;

Alignment Scores:
Pred. No.: 2.9e-09 Length: 1070
Score: 344.50 Matches: 243
Percent Similarity: 32.74% Conservative: 180
Best Local Similarity: 18.81% Mismatches: 429
Query Match: 1.77% Indels: 440
DB: Gaps: 60

US-10-087-013-1 (1-10628) x PVDR_PLAVS (1-1070)
QY 6918 ATGTTTCACAGATGGGACGAGAAATTTGTAATTAAGGAGAACGATTGTTAAAT 6977
DB 16 LeuLeuHisLysValAsnAsnValLeuLeu-GlnTyrGlnThrLeuGlnVal 34
QY 6978 GAGCGCGCGCTGAAGATATGAGTGAATGTAATGACGCTGAACACCAAGATG 7037
DB 35 -----CysLysAsnGlnTyrValLysGlnGlnAsnGlyTyrLysLeuAlaLys-- 50
QY 7038 TGCACAGCGCTGTGAACATATCAAAATTTATTAAGAGTGGAAACTGATATGAAAG 7097
DB 51 -GlnHisHisCysValGlnGlnAspAsn--LeuGlnLysTrp--Le 64
QY 7098 ACAGAGAGAAAGTTCAAAAGATTAAGATGCGCAAAAGATTAAGATTAATCTCTTAC 7157
DB 64 uGlnGlnThrAsnGlnLysArgSerGlnGlnAsnLysLysTyrLys--TyrGlnVal 83
QY 7158 TGAAGAGACATAGAGAGAGACATGCTCATGAATTTTAACATGAATTAAGA 7217
DB 83 rGlnLeuLysLysLysTyrAlaGlnMetAsnGlnLysArgSerSerArgLysLeuLysGln 103
QY 7218 A-----TTATGGGCAATTAAG----- 7233
DB 103 uSerLysThrGlnLysAlaHisAsnPheGlnGlnLysSerTyrMetGlnGlnLysAspGln 123
QY 7234 -GATTTCTTGTATGCAAAACCTTCTGACACATCCCAAAACACACACACATCACA 7292
DB 123 YAspLysThrGlnGlnGlnLysAspGlnLysHis-----LysThrAspSerLysThrAs 141
QY 7293 ATCATCCGATGCTATATATATGACAGATGCGTGATTAATGTTCTGAGAAATTAACA 7352
DB 141 pAsnGlnLysGlnLysAlaAsnAsnLeuValMetLeuAspTyr----- 154
QY 7353 GTGTGAGTGTCTGCAACTTTCACAAAGGATGATGATTCATACCAAAAAATTACTGA 7412

```


Db 155 -----GluThrSerSerAsnGlyGln----- 161
QY 7413 ACCTAAATAACCTATGATGTTGTAGAGAAAGCAGATATTATTAATCTAAAGAACAGCA 7472.
162 -----ProAlaGlyThrLeuAspAsnValLeuGluPheValThrGlyHisGluG 178
QY 7473 AATAATATGATATTAACCTTGAAGAAAAATTATACCTATTGAGTCTACAAAGAAAA 7532
178 yAsnSer----- 180
QY 7533 GGAATAATTAATAGTTGACATAATATATCT-----TGCATCTTAAGAAACCTTA 7586
181 -----ArgLysAsnSerSerAsnGlyLysAsnProTyrAspPheHisLysLysThrI 199
QY 7587 TGCACCT-----GATTAATATATAGAGAAAGAAACCCCTGTGAAATAGAGAGAA 7637
199 eSerSerAlaIleIleAsnHisAlaPheLeuGlnAsnThrValMetLysAsnGlyAsnTyr 219
QY 7638 AATTCCTTTAAGTATGATATATGAAATGGAATGTTACAAAAATTCAAGTTCTATACGA 7697
219 rLysAlaGlyLysArgArgGluArgAspTrpAspCys-----AsnThr 232
QY 7698 GAAAAAAGATATGTTGACCTCCAGAGAGAAACATATGCTTAAAGAAATTAGATGA 7757
232 rLysLysAspValCysIleProAspArgTyrGlnLeuGlyMetLysGluLeu----- 250
QY 7758 AATTAATTAATGAAGATTAGAGATAGTAT----- 7788
251 -ThrAsnLeuValAsnAsnThrAspPheHisArgAspIleThrPheArgLysLe 270
QY 7789 -TATTCCTTAATAATGTTGCTGCAACCTGCAAGAAATGAAGCAATAGACATAATAAAAA 7847
270 uTyrLeuLysArgLysLeuLeuLeuTyrAspAlaValGluGlyLys-----AspLeuLeuLeu 289
QY 7848 CTCAACCTCAGAGAGAGGCGCAATGAAATCCAAATATGATAGTACTAGTAATATAGTT 7907
289 sLeuAsn-----AsnTyrArgTyrAsnLysAspPheCysLysAspIleArgTrpSerLe 307
QY 7908 CGGTGATCTGGGTGACATAGTATAGAGAACAGATATGATGATGTTGCTTACTATACC 7967
307 uGlyAspPheGlyAspIleIleMetGlyThrAspMetGluGlyLys-----GlyTyrSerLys 326
QY 7968 TCCCGTATGAATTAATATATAGTTTGAATACATATATGAAAAATGAGAAATAA 8027
326 sValValGluAsnAsnLeuArgSerIlePhe----- 336
QY 8028 AATTAAGATGAGAAATTAATATACAGATGACATACGTTGCTTGTGCTGGGATGC 8087
337 -----GlyThrAspGlu-----LysAlaGlnGlnArgLysGlnTrpTrpAsnGln 352
QY 8088 TAATAGAAAGATATTTGGAAGCAATGACGTGCMAAGCAGCAAGAGATGCAAAACTTTT 8147
352 uSerLysAlaGlnIleIleThrAlaMetLysSerVal----- 365
QY 8148 TAGAAAAAGAGATGATGATTTGACAGCATTAATATCAAGATTAAGTGTGACA 8207
366 -----LysLysArgLeuLysGly----- 371
QY 8208 TAAGAGCATCCACCTGTGATATATATA----- 8238
372 -----AsnPheIleTrpIleCysLysLeuAsnValAlaValAs 384
QY 8239 -----CCTCAACGGTTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 8291
384 nIleGluProGlnIleTyrArgTrpIleArgGluTrpGlyArgAspTyrValSerGluLe 404
QY 8292 GATGGAAGATTTGGAAGAAATTTTAAAAATGATGATGATGATGATGATGATGATGATGATG 8351
404 uProThrGluValGlnLysLeuLysGluLysCysAspGly----- 417
QY 8352 CAAAGATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 8402
418 -LysIleAsnTyrThrAspLysLysValCysLysValProProCysGlnAsnAlaCysLys 437

QY 8403 AGAATATAAAAATTTTCTTAAATGGAATCTTATTCGATATACAAATCAATATAATA 8462
437 sSerTyrAspGlnTrpIleThrArgLysLysAsnGlnTrpAspValLeuSerAsnLysPhe 457
QY 8463 CAAAGATTTGTATGAAACCAATATATACAAAAATCTGACTTATGATCATCTTCAAAA 8522
457 e-----IleSerValLysAsnAlaGluLysValGlnThrAlaGln 470
QY 8523 TTTTGTACAAAAGTTTGAAGAACTTTTAAAGTAAATGTTCTGTGAGAGCTTTCTGAATA 8582
470 yIleValThrProTyrAspIleLeuLysGlnGlu-----LeuAspGluPheAsnGluVal 488
QY 8583 TCTTCATGAAACAAAGTAACTGTTGAATTAATTAATTAATGAAGAAATGATGTTCTTCAA 8642
488 lAlaPheGluAsn-----GluIleAsnLysArgAspGly----- 499
QY 8643 TATACGACATATGCTTTGCAAGAAACCAAAAAAGTTTAAAGAAAGCTTGCAAGTTGAC 8702
500 -----AlaTyrIleGluLeuGlyValGlySer 508
QY 8703 ACTACCTCTAAGATTCATTTGATATGCTTACCGATCAAAACAAAGATGATGTA 8762
508 rVal-----GluGluAlaLysLysAsnThrGln 517
QY 8763 GGAATTAACAACCTTTTACCTTCTGCTGAGAGAAATGATTAATTAATTAATTAATTAAT 8822
517 nGluValValThr-----AsnValAspAsnAla 526
QY 8823 GAAGCATACCTTCTTCTTAAATAGTTCA-----GATGATTAACAAG 8864
526 aAlaLysSerGlnAlaThrAsnSerAsnProIleSerGlnProValAspSerLysAla 546
QY 8865 TGTATGATCTCTCAAGAAAGATTAATGATGATGATGATGATGATGATGATGATGATGAT 8924
546 aGluLysValProGlyAspSerThrHis-----GlyAsnVal 558
QY 8925 TAGAAAAGTGAATTAAGAAATTTTAAAAAACTTCTTACTTCTGCTTCAAGTCAAG 8984
558 lAsnSerGlyGlnAspSerSerThrThrGlyLysAlaValThr-----GlyAspGln 575
QY 8985 ACAATGTTTGAAGCAAAATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9044
575 yGln-----AsnGlyAsnGlnThrProAlaGluSerAspVal----- 587
QY 9045 TAGTTATGACATTTATCCGATATTAATTAAGAAAGATGATGATGATGATGATGATGATGAT 9104
588 -----GlnArgSerAspIleAlaGluSerValSerAlaLysAsnValAspProGln 604
QY 9105 TGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9164
604 nLysSerValSerLys----- 609
QY 9165 GTGGGAAAAATATACAGCTCAGATATGCGACGCTATGTTATGTGATATTAATTTGCTAC 9224
609 ----- 609
QY 9225 TTCAAAAGTAACTTAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 9284
610 -----ArgSerAspAspThrAlaSer 616
QY 9285 GTTTCCTTCTGTTGTTAATGAATGGCGAAAGCAAGCATGTAAGAAAGAAACATGTAAG 9344
616 r-----ValThrGlyIleAlaGlnAlaGlyLysGluLysLeuLysAla 631
QY 9345 TGATTCATTTAAAAACAATAATGCTCTGTTCAACGAAGATTAATTTGAAGCGTCACAAAT 9404
631 rAsnSer-----ArgProSerGluSerThrValGlnLysAsn----- 643
QY 9405 ATTAAGCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9464
644 -----SerProGly-----AspAspThrValAsnSerAlaSerIleProValVal 658

OY	9465	AAAAATCAAGAAATAAT---CTAATAATATAATTAACCATTTAAAGATCAATCTTC	9521
Dd	659	-----SerGlygluasnProIeuValThrProTyrasnGlyleuAlaHisSerLysas	676
OY	9522	AGGTAAATATAGCAATAAACCATCTGAAGA-----AATGTCACCTCATATATAAATC	9575
Dd	676	pasnSerAspSeraspGlyProAlaGlnSerMetAlaasnProAspSerAsnSerLysGl	696
OY	9576	AAAAGATTCTCAATCGGCTTTGGAGTTAAATGATATTAATGAAATAGTTACAGGAACA	9635
Dd	696	yelutHrGelYlys-----GlyGlnasphaasnapeMalaLysalathLy	711
OY	9636	AAATATGAAATAATGAAATTCAGAAAGACTACAAAAAATATATCTCGGTATATATTT	9695
Dd	711	sAspSerSerAsnSerSer-----	717
OY	9696	TGTTGAGATGMAACACACAAAAATCATGTACTAGATGAAATATATAAAGAAAGACCA	9755
Dd	718	-----AspGlyThrSerSerAlathrGlyas	726
OY	9756	AACAGTTGCTCTAAGCACACTATATTTCTTACACCACCTAGATATCTTCTATCAAGC	9815
Dd	726	pThr-----ThrAspAlaValAsp-----	732
OY	9816	ACCTTATTTCCAAACACATGAGTAGACACAAATATGATCTTAAAAATGATATATTGAAAG	9875
Dd	732	-----	732
OY	9876	TAGATCTCTGTTGATATTGTATATGCAGGCTTAGTGTTGATAGCGCTTATTCATGAAGA	9935
Dd	732	-----	732
OY	9936	AAAATTAACAATCGTCTGTGCACCTTTGGCGTACTGAATATCCCGCAAGAGATATGG	9995
Dd	733	-----ArgLysLeasn-----LysGl	738
OY	9996	AATGCTACGTTGGATCCAAAATAGATACATACATATAGAGTGTCATTAAGG	10055
Dd	738	yAlaPro-----GlnasparGaspLysThrValGlySerLysAspGlyGlyGlu--	755
OY	10056	CMAAACATATATATATATGGAAGACATACTAGGAGATGAGAT	10105
Dd	756	-----AspnSerLiasnLysaspsAlaIatHrVala	767
OY	10102	-----AAATATATGGGACTTATCTTCCTCATATTTACTTCATCCGAAAGTGA	10155
Dd	767	LgylgluasparGlllearglunSerLiaGlylerThrnsnsparsGlyss--	786
OY	10155	TGAAGAAATGGATTAATGATATATATATGTCACAGTAGTCTTAATAATAAACATTGAT	10214
Dd	787	-AsnaspthrGllulYasnGlyalaserThrProksperLysGlnserGluaspaIath	806
OY	10215	AGAGTGTACTGAAACCATCAAAAAAGGATATACCAAGTATGATACACCAAGTAATGA	10274
Dd	806	rAlaleuSerLysThrGlnSerleugluserThrGlnSerGlyasparGthrThrnsnas	826
OY	10275	TACACACAGTACG-----AATAGATTTATGATGATANAAGAAATGACAGCAACCTGA	10328
Dd	826	pThrThrAsnSerleuglUasnLysasnGlyLysGlnLysaspleuGlnLynshsas	846
OY	10329	TTTTTGATCTCAATATTTCCAAATATACGAACCAATATATATATTACAAAAGTCAGATAT	10388
Dd	846	pPheLysSerAsnaspthrProasnGlnGlnProksnSerasp--GlnThrThrAspal	865
OY	10389	TCCAAGTAATACGAACCTTAATACCTTTATATCTCGATATATGCCGTAAGAAAAACCTTTAT	10448
Dd	865	aglu-----GlyHisAspArgdasSerellieLysasnbsprLysalacIunrGratYghshiswe	884
OY	10449	TATATCTATTCATGATAGGATTTATATCTGCGAAAGAAATATAGTTATATTTATATAT	10508
Dd	884	t-----AsnLysAspThrPheThrLysasnThrAnsSerethLshIsleunsbe	900
OY	10509	GAGTACATAATACAAATANGATATTTCCAAATAGCTAGAAATGATCTTATATGAGCTAT	10568

5980 AAAGAGTTGAAGGAAGTGAATTGTCAGTGCACAGAGTCCACACGCTGACGAGG 6039
563 LysGlnLysSerLysGlnSerLeu-----LysIleGlnSer 576
6040 GAAACACCGTCACCGAGGATACATGATACAAAGCGAGCGATCGAAGAAAGAGCG 6099
577 GlnThrValAsnGlnAlaLysGlnAlaIleThrLeuLysSerGlnLysMetAspLeu 596
6100 AAACAGACGGCGCTCAAAACAGCCCAAAAGTGAATAATTCACACAGAAATGCGA 6159
597 GlnSer-----ArgIleGlnLeuLysGlnLeuGln 608
6160 GCACAAACAGAAACCCGACGACGACACAAACAGAAACGAAACATGACAGACACA 6219
609 -----GlnLeuLysThrSerValProAsn 616
6220 ACAACAGAAATCTGACGCGGACAAATGGTAAAGCCATTCTTGCATTAACAGATAGC 6279
617 GlnAspAlaSerLysSerAsnValThrIleLysGlnLeuThrGlnThrLysAspLeu 636
6280 AGGGGTGATATGAGGGTGTATCCAAAAAGCTATGACAAATATCTTAATGGGTGT 6339
637 GlnSerGlnValGlnAsnLeuGlnThrArgIle-----SerGln 649
6340 ATTGTAGTGAAGTCTAAAGAAATGAAATGGCATGTATGCTCTAGAGAAAGAAA 6399
650 IleThrArgGlnSerThrGln----- 656
6400 TTATGTATTAATATATACAAATTTTAATATGAACT-----GAAATAAG 6447
657 -----AsnMetLeuLeuAsnLysGlnIleGlnAspLeuLysSerIle 672
6448 CGTGACAAATGATTAAGAGCGCTTTTATTAATGTGACAAATA-----GAAACTCA 6501
673 SerAspIleSerIleLysLeuGlnLysGlnLysSerSerArgIleLeuAlaGlnLys 692
6502 TTTTGTGGTAAATATATTAATGAAAACTCGACAGAAATGAAATGAAAAATGCA 6561
693 PheLysLeuLeuSerAsnThrLeuAspLeuThrLysAlaGlnAsn----- 707
6562 ACAATTCACGATGATTTAAAGAAATATGATATGATATGATGATGATTAAGATAG 6621
708 -----AspGlnLeuArgLysArgPheAspTyrLeu----- 717
6622 TTTTGGAACTGATATTTCTAATGATAAAAATTAATTAAGTAAACAAATAGTAA 6681
717 ----- 717
6682 ACCATTCATGAATAATTAAGAAAAACGATAAAAAAAGATGAAGATTCAGT 6741
718 -----GlnAsnThrIleLeuLysGlnAspSerLysThrIleGlnThrLeuAsn 733
6742 AAAATATTTTGGAGAAAAATTTATTTTGGAGAGAAATGATATGATGATTAAGT 6801
734 GlnThrValSerCysLysSerLysLeuSerIleValGlnThrGlnLeuLeuAsnLeu 752
6802 TATCATCTCAGACGAAACGAAAAAGAAAAATTAAGATTAATTAACAGTACAT 6858
753 -----LysGlnGlnLysLeuArgValHisLeuGlnLysAsnLeu 766
6859 -----GACATGACCAACTGACCGCT-----TCCCTTGAAGACTTTGTAAGAG 6903
767 LysGlnGlnLeuAsnLysLeuSerProGlnLysAspSerLeuArgIleMetValThrGln 786
6904 CCCCAGATTTTGAATG-----TTCAAGAGAGGCGAAGAAATTTTGTAAAT 6951
787 LysGlnThrLeuGlnLysGlnArgLysLeuGlnLeuGlnLysThrArgLysSerCysGln 806
6952 AAGAGAGAGAGAGAGT-----TTAAATTTGAGCGCGGCTGTAG 6993
807 LysLysIleAspGlnLeuGlnAspAlaLeuSerGlnLeuLysLysGlnThrSerGlnLys 826
6994 GAATAT-----GAGTGTATGCTAGTATGACGCGTAAAGACACAGAA 7035

827 AspHisHisIleLysGlnLeuGlnLysAspAsnAsnSerAsn----- 840
7036 TGTGAGAGCGCGTGTATACATATCAAAATTTTATTAAGAAAGTGAAGAAATGTA 7095
841 -----IleGlnThrPyrGlnAsnLysIleGlnAlaLeuLysLysAspTyrGln 856
7096 -----AGACAAAGAGAAAGTCAAAAGGATTAAGATGCGCAAAAGTATAGAT 7146
857 SerValIleThrSerValAspSerLysGlnThrAspIleGlnLysLeuGlnLys 875
7147 TATCCTTCTACGAAAGACATAGAGAAAGCAACATGCTGTCATGATTTTAACATG 7206
876 ValLysSerLeuGlnLysGlnIleGlnLysAspLysIleArgLeuHisThrTyrAsnVal 895
7207 AAATTAAGATTTATGTCAGATTAAGATGTTCTGTGTATGCAAAACCTTCTACAA 7266
896 MetAspGlnThrIle-----AsnAspSerLeuArgLysGlnLeuGlnLysSerLys 913
7267 CTACCAAAAACACACAAACATCACAAATCCGATGCTAATGATATGCGAATCGCTG 7326
914 IleAsnLeuThrAspAlaLysSerGlnLysGlnLysAspLeuThrGlnThr 933
7327 GATTAATGTTCCGAGAAATTTAACAAGTGTGCTGCTGAACTTTCAAAAAGGATCT 7386
934 SerGlnSerLeuGlnGlnThrAsnSer---LysLeuAspGln---SerPheLysAspPhe 951
7387 ATGATTCATACAAAAAATATCTGACCTTAATATACCTATGAAATGTGTAGAGAAACA 7446
952 ThrAsnGlnIleLysAsnLeuThrAspLysThrSerLeuGln-----AspLysIle 969
7447 GCATATTTTATATCTAAAGAACAGCAAAATATATGATATGATTAACCTGAGAAATTT 7506
970 SerLeuLeuLysGlnGlnMetPheAsnLeuAsnGlnLeuAspLeuGlnLysGly 989
7507 ATACCTATGAGTGTACAAAGAAAGAAAGTAAATATGTTGCACTAATTAATCTCT 7566
990 MetGlnLysGlnLysAlaAspPheLysArgIleSerIleLeuGlnAsnAsn----- 1007
7567 TGCATCTTAAGAAACCTTATGACCTGATTAATATATAGAGAAAGAAACCTTGTGAA 7626
1007 ----- 1007
7627 AATGAGAGAAAAATCGTTTAAGTAGATTAATGAAATGTAACAAATTCAAAG 7686
1008 AsnLysGlnValGlnAlaValLysSerGlnLysGlnLysLysSerLysIleGlnAs 1027
7687 TTCTATCAGAGAAAAAAGAGTATGTACCTCCACAGAGAGAAACATATG-----TG 7739
1027 AspLeuAspGlnGlnThrIleThrAlaAsnThrAlaGlnAsnAsnThrGlnGlnLys 1047
7740 CTTAAGCAATTTAGATGAATTAATA-----ATTGAAGACTTAAGATATGATATCTCT 7796
1047 GlnLysHisIleAspValSerLysThrIleSerGlnLeuArgGlnLeuHisThrTyr 1067
7797 AAAATGTTCTGTCGACCTGACCAATGAAAGATGACATATATAAAAACTTCAATC 7856
1067 TyrGlnGlnValLysThrLeu---AsnLeuSerArgAspGlnLeuLysAsnAlaLeu 1086
7857 AGAGAAC-----GGGTGCGCAATGAATCCAAATGTGATCTATGTAATATAG 7904
1086 GlnLysGlnLysSerThrPheSerGlnLysGlnSerLeuGlnGlnLeuAspLeu 1106
7905 TTTCGCTGATCTGGGTGACATAGTATGAGACAGATATGTTACGAATGGTGTACTT 7964
1106 ThrSerArgIleGlnLysPheSerGlnAsnLysLeuLys----- 1121
7965 ACCCTCCGTAGAAATTAATATATAAGTTTTGAATACATATATGAAATGCGAAA 8024
1122 -----AspGlnIleGlnIleThrAlaAlaAsp-----LysGlnValAsnAs 1136
8025 TAAAAATTAAGTGAATTAATATACAGATGACAAAGCTTTCGTTGCTTGGTGGGA 8084

```
Db 1136 nSerThrAsnGlyProGlyLeuAsnAsnIleLeuIleThrLeuArg----- 1151
QY 8085 TGGTAATAGAAAGATATTTGAAAGCAATGACGTGCAAAAGCAGAGATGCAAAACT 8144
Db 1152 ----ArgGluArgAspIleLeuAspThrLysValThrValAlaGluArgAspAlaLysMe 1170
QY 8145 TTTTACAAAAGAAAGATGATGGATTGAAACCATTAACATTA-----ATACA 8192
Db 1170 tLeuArg-----GlnLysIleSerLeuMetAspValGluLeuG 1183
QY 8193 AGATAGCTGTGCACATAGACCATCCACCTGTTGAT----- 8229
Db 1183 aaSPalaArgThrLysLeuAspAsnSerArgValGluLysGluAsnHisSerSerIleI 1203
QY 8230 -----GATTATATACCTCAACGCTTCGATGGATGAC----- 8262
Db 1203 eGlnGlnHisAspAspIleMetGluLysLeuAsnGlnLeuAsnLeuLeuArgGluSerAs 1223
QY 8263 -----GAAAGCTGTGAATATATTTGTAAGCCTGATGAGAGA 8300
Db 1223 nIleThrLeuArgAsnGluLeuGluAsnAsnAsnLysLysGluLeuGlnSerG 1243
QY 8301 ATTGAGAAATTTAAATAATCATGTGATCATGTGTAATAACATGACAGATGCAAGATGA 8360
Db 1243 uLeuAspLys-LeuLysGlnAsnValAlaProIle-----GluSerGluLeuT 1259
QY 8361 TTAATGATGAAATAAGTGTGACAGCTGTAAACGAGATGTCAGAGA--TATATAAATTTTG 8419
Db 1259 hAlaLeuLysTyrSerMetGlnGluLysGluGlnGluLeuLysLeuAlaLysGluLuv 1279
QY 8420 TTCTTAATGGAATCTCTATTGATATACAAATCAATTAATACAAAGATTTGATATAC 8479
Db 1279 aHisArgTrpLysLys-----ArgSerGlnAspIleLeuGlnLysHisGluG 1295
QY 8480 AACCAATATATACAAATACTCTATGATCATGTGTTCAAAATTTTGTCAAAAGTTGA 8539
Db 1295 In-----LeuSerSerAspTyrGlu----- 1302
QY 8540 AAACCTTTAAAGTGAATGTTCTGTGTGAGACCTTTCTGATATATCTTCATGAAACAATA 8599
Db 1303 -----LysLeuGluSerGluIleGluAsnLeuLysGluLuvLeuGluAsnLysGlu 1320
QY 8600 AGTGT-----TTCAATTAATTAATTTAATGAAATGATGCTTCCATATACGAA 8650
Db 1320 rGlnGluLysAlaGluLysLysPheAsnArgLeuArgArgAlaGlnGluArgL 1340
QY 8651 CATATGCTTTGCAAGAAACACCAAAAGTTATAAGAAAGCTGCACTGATACACTACT 8710
Db 1340 eUlySerThrSerLysLeuSerGlnAspSerLeuThrGluGlnValAsnSerLeuArgAsp 1360
QY 8711 CTAGAAATCCATTTGATTAATGTTCTTACCGATCAAAACAAAGATGATGTAAGCAATTTAC 8770
Db 1360 lAlaAsnValLeuGluAsnSerLeuSerGlnAlaAsnAla--ArgIleGluLeuG 1379
QY 8771 AAACCTTTACCTCTGTCGAAAGATATATATGATATATATCTGTAATTTGAAACCAT 8830
Db 1379 lAsnAlaLysValAlaGlnGluLysn-----AsnGlnLeuGlu--AlaI 1393
QY 8831 ACCTTGTTCTTAATAGTTCAAGATGATACAAAGCTGATGATGATCCCTCAAGAAAGAC 8890
Db 1393 lLeuArgLysLeuGlnGluAspAlaGluLysAlaSerArgLuvLeuGlnAlaLysLeuGlu 1413
QY 8891 ATTATATACACAGACCTATACCTGCAATATATATATACAAAAGGTGATTAAGAAATTTTAA 8950
Db 1413 lUeSerThrThrSerTyrGluSerThrIleAsnGluLysLeuAsnGluGluIleThrThrLeu 1433
QY 8951 AAAAAAACTTCTACTTTCGCTTTCAGTCAAGACAAATGTTAGTCAAAAATATATAAT 9010
Db 1433 ySgIuGluIleGluLysGlnArgGlnIleGlnGlnGlnLeuGlnAlaThrSerAlaAsnG 1453
QY 9011 CGGAAGAGAGTTG-----TGCCTTGAGGCAATGAATATAGTTAGTCAAGATATTTCCG 9064
Db 1453 lUglnAsnAspLeuSerAsnIleValGluSerMetLysLysSerPheGluGlu-----A 1471
QY 9065 ATATATATTAAGCAACTGATATGATGACACTTCATATCTGAAAAAT----- 9114
Db 1471 sPluLysIleLysPheIleLysGluLysThrGlnGluValAsnGluLysIleLeuGluAlaG 1491
QY 9115 --AAAAAATATTTGAAACATCAATGACACACCGAAATCTGTAACATGTTGGGAA 9172
Db 1491 lGlnArgLeuAsnGluProSerAsnIleAsnMetGluGluIleLysLysSerGluS 1511
QY 9173 ATATATACAGTCAGATATGGACGCTATGTTATGATATTAATATCTACTTCAAAAG 9232
Db 1511 eArgLuv-----HisGluGlnGluValSerGlnLysIleArgGluAlaGlu 1536
QY 9233 TAACATATGATGAGATGATGTCATATACCAAGATGAAAGAACTAATCAATGTTCTTC 9292
Db 1526 lValLeuLysLysArg--lLeuArgLeuProThrGlnGluLysIleAsnLysIleIleG 1545
QY 9293 GTTGCTTAATGAAATGGGCAACAGACATGTAAGAAAGAAACATGTAAGTATTCAT 9352
Db 1545 lUatGlyLysGluGluLeuGlnLysGluPheGluGluLys-----ValGluGluArgI 1563
QY 9353 TAAACAAATAATGTCCTGCTCAACAGAAATATTTGAAAGCTGCAATATATTAAGAC 9412
Db 1563 lElySer----- 1568
QY 9413 AACCTGATGTCAGATATATAGAAATATATATAGCTTGATATATATGTAATAAATA 9472
Db 1568 lInsArgLuv-----GluIleAspValValLeuArgLysG 1579
QY 9473 CAATGGA--AATCTAATTAATTAATTAAGCAATTAAGCAATTAATCTCAGATATA 9529
Db 1579 lUleuGluAlaLysValGlnGluLysGlnLysGluLeuGluAsnGluLysAsnLysLys 1599
QY 9530 TAGCAATTAACCAATCGAAGAAATGTCAGTCATATATTAATTAATTAATTAATCTCAAT 9589
Db 1599 eUglnGluGluLeuLysAspAlaProHisSerSerHisIle--SerAspAspAlaTyr 1618
QY 9590 GCGCTTGAGATTAAATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9649
Db 1618 sPluLysLeuArgAla-----GluIleGluSerArgLeuArgGluLuvPheAsn 1634
QY 9650 ATGAATTCAAAGAGTACTAAAAAATTAATCTCTGTTTATATTTGTTGTAACATGAA 9709
Db 1634 sGlnLeuGlnAlaIleLysLysLysSerPhe-----AspGluG 1647
QY 9710 CACACAAATAATCATGATCTAGATGCAAAATATA-----AAAGAAAG 9751
Db 1647 lYlysGlnGlnAlaMetLeuLysThrThrLeuLeuGluAlaGlyLeuAlaLysMetGluS 1667
QY 9752 AGCAAGAGTTGCTGTAAAGCACTATATTTCTTTACACCCCATGATGATTTCTTATC 9811
Db 1667 eArgLuvSerGluThrLysGlnSerAlaGluSerProProLysSerValAsnValG 1687
QY 9812 AAGCACTTTATTTCTCAACACATCAGAGACACAATATGAT-----CCTAAAAATG 9862
Db 1687 lAsnProLeuGlnLysLeuProArgLysIleGluGluAsnSerAsnSerPheAsn 1707
QY 9863 ATATATGAAAGATGATATCTCTGTTGTTATGATGCGCGTGTAGTTGATACGCTTC 9922
Db 1707 rOleuLeuSerGluLuv-----L 1713
QY 9923 ATTTCATGAGAAAAAATTCAAATGCTGTGACCTGTTGCGATATCTGATATCCCGC 9982
Db 1713 ySleuLeuLysLeuAsnSerLysSerSer-----S 1723
QY 9983 AAGGAGATGAGAAATGCGCTGTTGGAATGCAAAAATAGTACATA----- 10029
Db 1723 eArgLysIlePheAsnProPheThrSerProSerProAsnLysHisLeuGlnAsnAspAsn 1743
QY 10030 --CCATATAGAGTGTCCATATTAAGGCAAAACATATATATATGAGCAAGAGATACTA 10087
Db 1743 sPluLysArgLuvSerLeuAlaAsnLysThrAspProProThrHisLeuGluLuvProSerPhe 1763
```


[illegible]

Db	514	Asn	-----	-ValTysAsnValGlySerGlyVal	-----	522	
Oy	4942	CAAAACCAAGGTCCTATATAC	CTACAAATTAATCTGAAAGAAATTCAGTGAAGATGTCCTTTT			5001	
Db	523	GlnSerTyrAlaProSerSer	-----	AsnProIleAsnGlnAlaValLysSerSerSerGly		541	
Oy	5002	CCCTTCGTGCGACTAGCATATAT	GTTCCTCAT	-----	GCATTCG	5037	
Db	542	GlnGlyTyrValGlnGln	AspSerValIleHisArgSerValAsnGlnGlyGlyLysSer			561	
Oy	5038	GATGGCAATATATACAGATCC	AGAGTTAAAGATGAAAAAGCGTCCGAAAAAGTTGATG			5097	
Db	562	SerThrAsnGlnAlaAspPro	GlySerGlnProGlyGlyProAlaSerArgSerValAsp			581	
Oy	5098	GAATGGCGCCACGCAAGAGG	TGCACATTTTGGCTCAATCTACTACCAAGAAAAAAGAAAAA			5157	
Db	582	GlnLysAlaGlyValPro	AlaIleSerAlaGlnGlnGlnGlnLysAspLysValProAla			601	
Oy	5158	GAGAAAAATAAAAAGTCGG	ATGGCGCACAAATATCTTATAGAGGTCGCCCTGTAGTCT			5217	
Db	602	GlnLAlaIlaIlaIleThr	GlnSerAlaValAlaProHisSerAlaAspLysThrProIle	ThrAla		621	
Oy	5218	ATGAAATATATAGTTTATG	ATTATTAAGCATATATCTAGTATTCATTAATTTGACAGAT			5277	
Db	622	-----	ThcGlnGlu			624	
Oy	5278	GAAAACCAAAAGACCGAG	AAAAATTTTGAACAAATATATTTAACAAAAATGCAACATCACTT			5337	
Db	625	AsnLysGlnAspArgThrGln	-----	ValAspGlyValAlaGlyGlyAspGlyLysAlaPro		642	
Oy	5338	GGCAAA	-----	GGAAGGATATGATCTACAGGAAT		5367	
Db	643	GlyProThrValSerSer	AspValProSerValGlyGlyLysAspSerGlyProSerThr			662	
Oy	5368	CCCCGATAGTACTGGCG	CAAAATTTTCTCGAACGAAATTAAGCAATGCTGTGTGCAACGCA			5427	
Db	663	ProAlaSer	-----	HisLeuAlaGlyGlnAsnGlyGln	-----	ValHisAsnGly	677
Oy	5428	ATGATATGCGGGTACAA	AGCTGTTAGGAGATGATGCAAAATAGTGGAAATATGTCCAAAGAACT			5487	
Db	678	-----	ThrAspThrGlnProLysGlnAspGlyGlnLysAlaAspPro			691	
Oy	5488	GATGAGATCTTAAAAAA	ATGGTGCTGTGACCTTCAGATGAT	-----	CATATACCTTGG	5541	
Db	692	GlnLysAsnIleGlyValLys	GlyLysGlnAspThrAspAspArgSerGlnLysSerLeu			711	
Oy	5542	GCGAAAAATCCGATGA	AGGTAAGTACTGCGTATCACTTCTTCGATGGTTTCCGAAATGGGT			5601	
Db	712	GlyProHisThrAspGlu	Arg	-----	AlaSerLeuGly	722	
Oy	5602	GAAGATTTTTCAAAC	ACTTAAGAAAAAGATTTGGAGAAATTTGGTAGGGCGGT	-----		5655	
Db	723	Glu	-----	ThrHisMetGlnLysAspThrGlnThrThrGlyGlySerThrLeuThr		739	
Oy	5656	-----	AATGATTTACTTGTGGTGAATGAAGAT	-----	AAAAAGAAAGAA	5697	
Db	740	ProGlnGlnAsnValSer	ValAlaSerAspAsnGlyAsnValProGlySerGlyLysAsnLys			759	
Oy	5698	TGTACAGATCGCGTGA	CAACATATAAAAAATTATTAGTGAGTGGAAACCAACATATGAA			5753	
Db	760	GlnAsnGlnGlyAlaThr	AlaLeu	-----		767	
Oy	5758	AAACAAATCAAAAATAT	GTGGAGAAATTAAGCAAAATATATTCGCGAGCATCTGTGCA			5811	
Db	768	-----	SerGlyAlaGlnSerLeuGlnSerGlnSerValHis			780	
Oy	5818	AAAGATGACAGAGACG	CTCGCGAATATTTAGACAAACAAATTAAAAAAATTTGTGAAAT			5877	
Db	781	LysThrIleAspAsnThrThr	HisGlyLeu	-----	GlnAsn	792	
Oy	5878	AAAAGTGAAGTTGTGA	ATTAATTAAGTATTAAGATATGTCACACACAGGATTAACGAT			5937	
Db	793	LysAsnGly	-----			795	


```

Db 624 -----|||||-----|||||-----
               -ThrGluGlu-----AsnLysGluGlyThrGlnMetAsp 634
QY 5332 -----|||||-----TCA 5334
Db 635 GYAlaIaIaGlyLysAspGlyLysAlaProGlyProThrValSerSerAspValProSer 654
QY 5335 GTTGGCAAGAGAGTACTACTACTAGCAAGAAATCCCGGTAGTACTGCGCGCAAAATTTTTC 5394
               |||||-----GlyLysAspSer-----GlyProSerThrSerAlaSerHisAlaLeu 670
Db 655 ValGly-----GlyLysAspSer-----
QY 5395 TGGACGAAAAATAAGAAATGTGTGTGTGAAGCAATGATATGCGGTGACAAACGTGTAG 5454
               |||||-----ValHisAsnGly-----ThrAspThr 683
Db 671 AlAGlyGluAsnGlyGlu-----ValHisAsnGly-----
QY 5455 GATGATGCAAGATAGTCAAAATAGTCAAGAGATGATGATCAAAATAAAATGCGTTC 5514
               |||||-----GlyLysAspSer-----GlyLysAspSer-----
Db 684 GluProGlySerGluAspGlyLysAlaAspProGlnLysAspGlyLysValLysGlyLys 703
QY 5515 GTACCTTCAGATGAT-----GATTATCCTATGCGGCAAAATCGCATGAGGTACTGCG 5568
               |||||-----GlyLysAspSer-----GlyLysAspSer-----
Db 704 GluAspThrAspAspArgSerGlnGlySerGlnGlyProHisThrAspGluArg----- 721
QY 5569 TATCAGTTCTCTGATGTTGCGCAATGGCGGATGTTTTCGCAACATATAAGAAAG 5628
               |||||-----AlaThrLeuGlyGlu-----ThrHisMetGlyLys 731
Db 722 -----|||||-----
QY 5629 GAATTCGAGAAATGTTGAGGGCGGT-----AATGATTATCTTGTGCT 5673
               |||||-----
Db 732 AspThrGluThrAlaGlySerThrLeuThrProGlnGlnAsnValSerValAlaSer 751
QY 5674 GATTAATGAGAT-----AAAAAGAAATATGACAGATCGCTGTACAAATATAA 5724
               |||||-----
Db 752 AspAsnGlyLysAsnValProGlySerGlyLysAsnGlyLysAlaThrAlaLeu----- 770
QY 5725 AATTTATTAGTAGAGTGAAGACACAGATATGAAAAAATCAATCAAAATATGATGAGAT 5784
               |||||-----
Db 771 -----
QY 5785 AAAGCAAAATATATCCAGCATCTCTGCGCAAAAGATGACAGAGACCTCGCGAATAT 5844
               |||||-----
Db 773 AlaGluSerLeuLysSerAsnGluSerValHisLysThrHisAspAsnThrThrHisGly 792
QY 5845 TTAGACAAACAATTAATAAAATTTGTGAATAAAGAGATTTGTAATATATAGTGT 5904
               |||||-----
Db 793 Leu-----GluAsnLysAsnGlyGlyAsnGlu----- 801
QY 5905 ATGAAGATGTGTCCACACAGGATTAAGTGTATGTAATATGCAAAATGTCGCCCATCA 5964
               |||||-----
Db 802 ---LysAspPheGlnHisHisAspPheMetLysAsnAspMetLeuAsnAspGlnHisSer 820
QY 5965 TTAGACGATGACCAAAAGAACTGAAGAAAGTGAATTTGCAAGTCCACAGAGTCCA 6024
               |||||-----
Db 821 SerAspHis-----ThrSer-----SerAspGlnThrSerSerA 832
QY 6025 CCACGTGTACGAAAGGAAACACCGTACACACGGTATCATCATATCAAAAGCGACGCA 6084
               |||||-----
Db 832 spHisThrSerSerAspGlnThrSerSerAspHisThrSerSerAspHisThrSerSerA 852
QY 6085 TCGAAAAAGAAAGCAAAACAGCGCGCTTACAAACAGCGCAAAAGATGCA-----A 6138
               |||||-----
Db 852 spGlnThrSerSerAspGlnThrSerSerAspGlnThrHisLysThrGlnGlyHisHisA 872
QY 6139 AATCAACACGAAATGCGAGCAACAAACAGCAAGCCACAGCAAGCGCAACAAACAGCA 6198
               |||||-----
Db 872 rGAspAsnValrGAsnProGlnLysLysSerGluAspMetSerLysGlyAspPheM 892
QY 6199 AAACGACATCAACAGCAAC-----AACACAGAAATCT--- 6231
               |||||-----
Db 892 eTArgAsnSerAsnSerAsnGlnLysLeuTyrSerHisAsnAsnLeuAsnAsnArg-LysLeu 911
QY 6232 -----GACGTG---GGCACAAATGTAAGGCGCATTTCTTTCG 6264
               |||||-----

```

```

Db 912 AsnArgAspGlnThrGlnHisArgAspValLysAlaThrArgGlnLysIleIleLeuMet 931
QY 6265 -----AATTAACACATACACAGCGGTGAAATAGAGCGTTGTAATCCAAAAACGTAT 6315
               |||||-----
Db 932 SerGluValAsnLysCysAsnAsnArgAlaSerValLysTyrCysAsn----- 947
QY 6316 GGACATATCTTAATATGGGGTTGTATTGTAGTAAGTCTTAAGAAAAATCA-----AAT 6369
               |||||-----
Db 948 -----ThrIleGluAspArgMetLeuSer 955
QY 6370 GGCATATGATGCGCTCTAGAGAAAAAAATTA-----TGATA 6408
               |||||-----
Db 956 SerThrCysSerArgGlnLysAsnLeuGlyAsnLeuGlySerIleSerAspPheCysLeu 975
QY 6409 AATTAATATCAATATTTAATATATGAACTGAAATTAACCGTACAAATGATTAAGAG 6468
               |||||-----
Db 976 AsnTyrPheGlnLeuTyrSerTyrGlnPheTyrAsnCysMetLysGluPheGluAsp 995
QY 6469 GCTTTATTAATATG-----GACGCAATAGAA 6495
               |||||-----
Db 996 ProSerTyrGluCysPheThrLysGlySerThrGlyIleValTyrPheAlaThrGly 1015
QY 6496 ACTCAATTTTGTGTTAAATATATTAATGAAAT-----CCTGCAGACAGAAATGATTTG 6552
               |||||-----
Db 1016 GYAlaPheLeuIleIleLeuLeuLeuPheAlaSerTyrAsnAlaAlaSerAsnAspTyr 1035
QY 6553 CAAAAATGGAACAATTCACAGATGATTT 6579
               |||||-----
Db 1036 GluGluGluAlaThrPheAspGluPhe 1044

RESULT 11
MS2_DICDI
ID MY22_DICDI STANDARD; PRT: 2116 AA.
AC P08799;
DE 01-NOV-1988 (Rel. 09, Created)
DI 01-OCT-1989 (Rel. 12, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MHCN.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092266; PubMed=3540939;
RA Warwick R.M., de Lozanne A., Leitwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
RL Dictyostelium discoideum."
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RX STRAIN=AX2;
RC MEDLINE=90353583; PubMed=2387408;
RA Luck-Vielmeier D., Schleicher M., Grabatlin B., Wippler J.,
RX Gettsch G.;
RT "Replacement of threonine residues by serine and alanine in a
RL phosphorylatable heavy chain fragment of Dictyostelium myosin II."
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel U., Gettsch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
RL Dictyostelium myosin heavy chain."
RN [4]
RP FEBS Lett. 227:71-75(1986).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Holden H.M.,
RT Rayment I.;
RL "X-ray structures of the myosin motor domain of Dictyostelium
RL discoideum complexed with MgADP. Bepx and MgADP.Alp4-";
RL Biochemistry 34:8960-8972(1995).

```


Db 161 ValAlaItyArgSerMetLeuAspasp---ArgGlnAsnGlnSerLeuLeuIleThrgly 179
1060 CATAATGACATGATTTGACACACTATTGTCAGAAAAAAGATTTGTCAT----- 1110
Db 180 GluSerGlyAlaGlyLysThrgLysnThr---LysLysValIleGlnItyLeuAlaSer 198
1111 -----TTGATATATAGTGTACTAGCTGT 1134
Db 199 ValAlaGlyArgAsnGlnAlaAsnGlySerGlyValIleGlnGlnIleLeuGlnAla 218
1135 TCGACTAATGCAAGTTTGGAACTTTGGTAGGATTCACAAAGACATTAAAAA 1194
Db 219 AsnProIleLeuGlnAlaPhe-----GlyAsnAlaLysThrThrArgAsnAsn 234
1195 CAAAAAGAAAAATATGAAAA----- 1215
Db 225 AsnSerSerArgPheGlyLysPheIleGlnIleGlnPheAsnAlaGlyPheIleSer 254
1216 -----GAAATACATCATTTATTATCGAACGTAACAAATTTGCAATATATTAAGT 1269
255 GlyAlaSerIleGlnSerItyLeuGlnLysSerArgValValPheGlnSerGluThr 274
1270 GAA---TATTATTAACAATTTATGAAAACTT-----AAGGA 1305
Db 275 GluArgAsnItyHisIlePheItyGlnLeuLeuAlaGlyAlaThrAlaGlnLysLys 294
1306 ACGCAATATGCACTAATGACACTTTTAAATTTACTAATTAAGAAAGAAATTTGTA 1365
Db 295 AlaLeuHisLeuAlaGlyProGlnSerPheAsnItyLeuAsnGlnSerGlyItyValAsp 314
1366 ---GGAGATTACACAGGAAAAAGATTTACTTTACTACACGCTGATGACAAAGG 1422
Db 315 IleLysGlnIleSerAspSerGlnLysPheLysIleThrArgGlnAlaMetSerPheVal 334
1423 ATATTTCGTCGATGATTTGCGCAAGTGTCCGCTGCGG-----GTCAATGT 1476
Db 335 GlyPheSerGlnGlnGlnMetSerIlePheLysIleIleAlaGlyIleLeuHisLeu 354
1477 GATGATATAAATACACACAAATCAGAT-----AATGATCGTGAAGT 1521
Db 355 GlyAsnIleLysPheGlnLysGlyAlaGlyLysAlaValLeuLysAspItyThrAla 374
1522 GTAAATATGAAGCTAATTAACCTCAGGCTGGAAGCTACTAATATCTCTC--- 1578
Db 375 LeuAsnAlaAla---SerThrValPheGlyValAsnProSerValLeuLysAla 392
1579 -----CTTTATAGTGTATGTAACAAAGTGAATATTAACAAATTAAGA 1623
393 LeuMetGlnProArgIleLeuAlaGlyArgAspLeu---ValAlaGlnHisLeuAsn 410
1624 AATTTTGTAACTGCTCACTAAT----- 1647
Db 411 ValGlnLysSerSerSerArgAspAlaLeuValLysAlaLeuItyGlyArgLeuPhe 430
1648 -----TCAAAAGATAAAAATAT-----CAAAATGGGATGCTATAT--- 1686
Db 431 LeuTyrLeuValLysLysIleAsnAsnValLeuGlnGlnItyArgLysAlaItyPheIle 450
1687 -----AAGATGAAATATATAAGATGTAACCTGAGCAAAATCTGAA 1731
Db 451 GlyValLeuAspIleSerGlyPheGlnIlePheLysValAsnSerPheGlnLeuLys 470
1732 ATCAATATGATTAATCCTAAGATATATATCTT-----CATATTTTTCATTAATG 1785
Db 471 IleAsnItyThrAsnGlnLysLeuGlnGlnIlePhePheAsnHisIleMerPheLysLeuGln 490
1786 GTTACATATTTATTAAGGATACTATTAAGTGAATGACAAACTTAACCTGTATA--- 1842
Db 491 GlnGlnGlnItyLeuLysGlnLysIleAsnItyPhe-----ThrPheIleAsp 505
1843 -----AATTAATNCAACGACGATGTATGATGATGAATGAACAAATTCCTTA 1890

Db 506 PheGlyLeuAspSerGlnAlaThrIleLeuAspLeuIleAsp----- 518
1891 TGTTTGACAGATGGGTTAAACAAAGAAAGAAATGATGATTAAGAACTGTTC 1950
Db 519 -----GlyArgGlnProProGlyItyLeuAlaLeuLeu 529
1951 ACAAAGAAAAAGATATACAGCAATCGTATTAATTAAT-----AATATCTTTT 2004
Db 530 Asp-----GlnGlnSerValPheProAsnAlaThrAspAsnThrLeuIle 544
2005 GAAGCTTATTTTAAAGTTATGATTAACCTGACCAAGATGAGCAAAAGAAAGAA 2064
Db 545 Thr-----LysLeuHisSerHisPheSerLysLysAsnAlaLysItyGln 560
2065 CTTATGAAATATTAATAAGAAAAAAGAAATGAGTTTCCAT----- 2106
Db 561 -----ProArgPheSerLysThrGlnPheGlyValThrHisItyAlaGln 576
2107 -----TTGAAATTAATAGGACTATTATGAGAAATCAATA 2142
Db 577 ValMetItyGlnIleGlnAspItyPheGlnLysAsnLysAspProLeuGlnAspLeu 596
2143 GAAGCTTGTATGATCAGCTTAAGAAAGAACTGCCAGATATGTAAGACAAATTAACAAC 2202
Db 597 GluLeuCysPheLysAspSerSerAspAsnValValThrLysLeuPheAsnProAsn 616
2203 GAAGCATGTGAACATCCCATTAATCA----- 2229
Db 617 IleAlaSerArgAlaLysLysGlyAlaAsnPheIleThrValAlaAlaGlnItyLysGln 636
2230 -----ACAAACAAACCG-----TGCTT 2247
Db 637 GlnLeuAlaSerLeuMetAlaThrLeuGlnItyThrAsnProHisPheValArgLysIle 656
2248 AAACCTCGTGGAGGACGACCAACCACTAAATAATTAAGAAATA----- 2292
Db 657 IleProAsnAsnLysGlnLeuProAlaLysLeuGlnAspLysValValLeuAspGlnLeu 676
2292 ----- 2292
Db 677 ArgCysAsnGlyValLeuGlnItyLeuArgIleThrArgLysGlyPheProAsnArgIle 696
2293 -----GCACATATCTTTAAAG----- 2310
Db 697 IleTyrAlaAspPheValLysArgItyTyrItyLeuLeuAlaProAsnValProArgAspAla 716
2310 ----- 2310
Db 717 GluAspSerGlnLysAlaThrAspAlaValLeuLysHisLeuAsnIleAspProGln 736
2311 -----AGTGCATACGAGAA 2325
Db 737 TyrArgPheGlyIleThrLysIlePhePheArgAlaGlnGlnLeuAlaArgIleGln 756
2326 GCACGAATTCGGTCTCTCATTAATTTGAAGGAAGGACACGAAGGTATATTAACGT 2385
Db 757 AlaArgGlnGlnArgIleSerGlnIleIle---LysAlaIleGlnAlaAlaThrArgGly 775
2386 GGGGTAGAGAAAGACTTCACAGACAAT-----TTATGATGATTAAG 2430
Db 776 TrpIleAlaItyGlyValItyLysGlnAlaArgGlnHisIleThrValAlaAlaArgIleIle 795
2431 ATAAACATTTCTAATCTGATCTGTTTTCAAATGACCATGTGATGCAAGACACA 2490
Db 796 GlnGlnAsnLeuArgAlaItyIleAspPheLysSerItyPro----- 809
2491 GGTGATGTATACAAACAAGATTGTGCTAGGACTGATG-----GAA 2535
Db 810 -----TyrTrpLysLeuPheSerLys 816
2536 GTGATCGGGAACACATGCGTAAAGATCAGCAAGATGTTATTAATCCCTCAAGACAGA 2595
Db 817 AlaArgProLeuLeuLysArgArgAsnPheGlnLysGlnItyLysGlnLysGlnArgGln 836

Db	1415	Thralalyslysvalllysallagluylulysalameetylslsyalagluylalaspetylalargser	1434
Qy	4705	ACG-----AATCCAAACAGATTACTTGAACAGGAATTTACTGCTAAGTTGT	4749
Db	1435	ThrlvsSerGluleuAspAlaSerplAlaLysAsnValSerSerGlugIntyr-----	1450
Qy	4750	GGTGATTAAGCCTGGAAAGTCCCTCGCTGGTACAAAGAAATATATACATTTGTGAA-----AAA	4806
Db	1451	-----ValGlnlleLysargleuAsnGluGluLeuSerGluLeuarg	1464
Qy	4807	CAGCCTACTATGATCCGCAACAACTTGTGGGTGCACAAAATTTATTTAGAAATGACGAC	4866
Db	1465	SerValleuGluGluAlaAspLuarGlySAsnSerIala-----IleLysAla	1480
Qy	4867	AAATFACTACATTTCGAGTAAACATTAAGTCCAAAGATTAGTAAAGAGCAACACA	4926
Db	1481	LysLysThrIaGluSerIalaLeuSerIeuleuLysAspGluIleAspAlaAlaAsnAsn	1500
Qy	4927	GGTGCTATTATGCGCAAAACAAAGGTCCTTAATACATACATTAAGTGAAGAAAGTACT	4986
Db	1501	AlaLysAlaLysAlaGluLuarGlySerLysGluLeuGluValArgValAlaGluLeuGlu	1520
Qy	4987	GAAAGATGTCCTTTCTCTCTCTGCTGACTACGTATATGTTT-----	5028
Db	1521	GluSerLeuGluAspLysSerGlyThrValAsnValGluPheIleArgLysLysAspAla	1540
Qy	5029	-----CATCATTTGATGGCAATTTATACAGATCCAGAACATTTAA-----	5067
Db	1541	GluIleAspAspLeuArgLuarGluAspArgGluThrGluSerArgLleLysSerAsp	1560
Qy	5068	---GATGAAATATGGGTGGCGGAAAGATTGATGGAATGGCGCGCAACGAGAGGCTCAAT	5124
Db	1561	GluAspLysLysAsnThrArgLysGlnPheAlaAspLeuGluAla-----	1575
Qy	5125	TTGGCTCATCTACNCAAGAAAAAGAAAAAGAAAAATAAAACTCGATGGCGCAC	5184
Db	1576	-----LysValGluGluAlaGlnArgGluValThrIleAspArgLeu	1590
Qy	5185	AAATATTCTTATAGAGTCCCGCCTGTAGTGCATGAATATAGTTTTATGATTTAAGA	5244
Db	1591	LysLysLysLeuLuu-----SerAlaIlelle	1599
Qy	5245	GATATATTTCTAGCTATGTATTAATTGGGAAGATGAAAMAAACAAAGACGAGAAATTTGG	5304
Db	1600	AspLeuSerThrGlnLeuAspThrGluThrLysSerArgLleLysLleGluLysSerLys	1619
Qy	5305	AAAGAAATATTTAACAAGAAATGAGACATCAGTTGGCAAGAGAGTACATGACTACAGCA	5364
Db	1620	LysLysLeu-----GluGlnThrLeuAla	1627
Qy	5365	AATCCCGGTAGTACTCGCGCAAAATTTTCTGGACAGCAAAATTAAGAAATGTGTGTGAAC	5424
Db	1628	Glu-----	1628
Qy	5425	GCAATGATATCGGGGTACCAACCTGATGGAGATGATGAATATGGAATAAGTGCACGA	5484
Db	1629	-----ArgArgAlaIaGluGluLysSerLysAla	1639
Qy	5485	AGTGATGAAGATCTAAGAAAAATGTGTTCTGTACCTTCAGATGATGATTTATCCTATGGGG	5544
Db	1640	AlaAspLugluIleArgLys-----GlnValTyr	1649
Qy	5545	AAAAATGCCATGAAGTACTCGCTATCAGTTTCTTCGATGGGTTTGGCGAATGGGGTGAA	5604
Db	1650	GlnIuValAspGlu-----LeuArgIaGlnLeuAspSerGluArg	1663
Qy	5605	GATTTTTCGAACATTAAGAAAGAAATTTGGAGAAATTTGGTAGGGGCGGTATGATATAT	5664
Db	1664	AlaAlaLeuAsnAlaSerGluLysLysLleLysSerLeuVal-----	1677
Qy	5665	ACTTGCTGTAAATGAAGATTAAGAAAGAAATGTACAGATCGCTGACCAATATATAA	5724

Db 1678 ----AlaGluValAspGluValLysGluLeuGluLysAspGluIleLeuAlaLysAsp 1695

Qy 5725 AAATTATTAGTACTGCGAACCACGATGATGAAAAACAATCAAAAAATATGCGACAAAT 5784

Db 1696 LysLeuValLysAla---LysArgAlaLeuGluValGluLeuGlu-----GluVal 1711

Qy 5785 AAAGCAAAATATATTCGAGACATCCTGTGGCAAAAGATGACAGAGCGCTGGCAAAAT 5844

Db 1712 ArgAspGluLeuGluGluGluLysSerArgSerArgSerGluLeuGluAspSerLysArgArg 1731

Qy 5845 TTAGCAAAACATTTAAAAAAATTTTGTAATAATATAAAGTCGAGATTGTGAATATTAAGTGT 5904

Db 1732 LeuThrThrGluValaGluAspLleLysLysLysTyrAspAlaGluValaGluLysnThr 1751

Qy 5905 ATGAAAGATGTGTCCACACAGCGATTAATCGTTGTGTATATGACAAATATGCCCATCA 5964

Db 1752 LysLeuAspGluValaLysLysLysLeuThrAsp---AspValAspThrLeuLysLysGlu 1770

Qy 5965 TTAGACGATGAAACCAAAAGAGTGTGAAGAAAGTATTTGTCAAGTGCACGAGGTCCA 6024

Db 1771 LeuGluAspGluLysLysLys----- 1777

Qy 6025 CCACGCTGACGAAGGGAACACCGTCACACGCGGTATCATGTATTCAAACGACGCA 6084

Db 1778 -----LeuAsnGluSerGluArg 1783

Qy 6085 TCGAAAAAAGACGGAACACAGCCGCCCTACAAAAACACCGCAAAAAATGTGAATAATCTA 6144

Db 1784 AlaLysLysArgLeuLysSer-----GluAsnGluAspPhe 1795

Qy 6145 ACACACGAATCGAGCACAAACACGACCCGACGACGACACCAACCAACGCAAAAAACA 6204

Db 1796 LeuAlaLysLeuAspAlaGluValLysAsnArgSerArgAlaGluLysAspArgLysLys 1815

Qy 6205 -----ACATCAACACCAACAACA 6225

Db 1816 TyrGluLysAspLeuLysAspThrLysTyrLysLeuAsnAspGluAlaIleAlaThrLysThr 1835

Qy 6226 GAATCTGACGTGGGCACCAATGCTAAAGGCCCATTTCTTCGAATTAACCAATAGCAGGGGT 6285

Db 1836 GluThrGluIleGluLysAla-----AlaLysLeuGluAspGluIleAspGlu----- 1850

Qy 6286 GGAATGAGAGGTGTATCAACCAAAACGATGAGCAATATCCTTAATGGGTTGTATTGTA 6345

Db 1851 -----LeuArg 1852

Qy 6346 GGTAACTCAACAAAAATGCAAAATGCCATATGTATGCTCCTCAAGCAAAAAATATATGT 6405

Db 1853 SerLysLeuGluLysGluGluLysAlaLysAlaThrGluAlaAspLysSerLysThrLeu 1872

Qy 6406 ATAAATTAATTAACAATTTAAATTTAATTTGAAGCACTGAAAAATTAAGCGTGACATGATATAAA 6465

Db 1873 GluGluLysIleAspAsnLeuArgAlaGluIleGluAsp-----GluGlyLysIleLys 1890

Qy 6466 GAGCGTTTATTAA---TGTGCAGCAATAGGAACATCAATTTTGTGTGTAATAATATA 6522

Db 1891 MetArgLeuGluLysGluLysArgAlaLeuGluGluGluLeuGluGluLeuAspArgGluThr 1910

Qy 6523 ATTGAAATATCCTGCAGCAGCAAAATGAATTCGAAATGGAACAATTCACATGCAATTTTAA 6582

Db 1911 ValGluGlu-----AlaGluAspSerLysSerGluAla-----GluGlnSerLys 1925

Qy 6583 AGAATTAATGTATTATACATATGCTGATTTAATAAAGATATGTTTTTTGCACTGATATTCTT 6642

Db 1926 ArgLeuValaGluLeuGluLeuGluAspAlaArg-----ArgAsnLeuGlu 1940

Qy 6643 AATGATAAAAAATTTATTAACGTGAACAATAATGTGTAACAACCATCTCAATGAAAAATAT 6702

Db 1941 LysGluIleAspAlaLysGluIleAlaGluAspAlaLysSerAsnLeuGluAlaGluIle 1960

Qy 6703 AAGCAAAAAACGATTAATAAAAAAGATGAAGCAATTAACGTAAATAATTTTGGCAGAAAAAT 6762

Db 1961 ValGluAlaLysGlyArgLeuGluGluGluSerIleAlaArgThrAsnSerAspArgSer 1980

Db 138 -----LysIleGluGluLeuThrAsnGlnLysSerPheMetLysGluLeuAsp 154
 QY 1960 -----AAGATATATACGCAATCCTATTATAGTATATATATATATCTT-----TTGAGGT 2010
 Db 155 SerThrLysAspLeuAsnTPaspLeuGluSerLysLeuThrAsnLeuSerMetGluCys 174
 QY 2011 TATTTTAAAGTTATGATTAACCTTGACAAAGATGACCAAAATGAAAGACTTATG 2070
 Db 175 ArgGlnLeuLysGluLeuLysLys-----LysThrGlnLysSerThrPasnAspGluLys 192
 QY 2071 GAAATATATAAAGAAAAAATGAGCTTTTCCAAATTGGAATAATATAGGACTATTTA 2130
 Db 193 GluSerLeuLysLeuLeuLysThrAspLeuGluIleLeuThrLeuThrLysAsnGlyMet 212
 QY 2131 GAGATGCAATAGAACTCTGTAGATTAAGAAAGAACTGCCAGATATGTAAGAC 2190
 Db 213 GluAsnAspLeuSerSerGlnLysLeuHisTyrAspLysGluIleSerGluLeuLysGlu 232
 QY 2191 -----AATATACAAACGAGCATGTGAACATCCCATTAATGCAACAC 2235
 Db 233 ArgIleLeuAspLeuAsnAsnGlnAsnAspArgLeuLeuIleSerValSerAspLeuThr 252
 QY 2236 AACCCGCTGTAACTCGTGGAGGACGACCCACTTAAATATATTAAGAAATAGCA 2295
 Db 253 SerGluIleAsnSerLeuGln-----SerAsnArgThrGluArgGlyLysIleGlnLys 270
 QY 2296 CAATACTTAAAGAGATGCAATCGAGAGAACGCAAAATCGTGCTTCATTAATGAAA 2355
 Db 271 Gln-----LeuAspAspAlaLys--AlaSerIleSerSerLeuLys 283
 QY 2356 GGAAGGACACAGAGATATATTAACGTGGGGTAGAGAAAGACATTCACAGCAAT 2415
 Db 284 ArgLysValGlnLysLysTyrGln----- 292
 QY 2416 TATATGATATATGATTAACATTCATCTGTTTCAATGAGCACATCT 2475
 Db 293 -----LysGlnHisThrSerAspThrThrValThrSerAspPro--- 305
 QY 2476 GATGCAAGGACACAGATGTGATATACAAACAGATTTGCTGAGAACTGAATGGAA 2535
 Db 306 AspSerGlnGlyThrThrSer----- 312
 QY 2536 GTGATCCGGAACATGCGTAAGATCACAGAGATGTTATATGCTCTAGAAAGACA 2595
 Db 313 -----GluGlnAspIlePheAspIleValIle----- 321
 QY 2596 CATATATGATCCAAATTTGGAACATTTTACAAACGATGATCACCCACTTAATGCTAAT 2655
 Db 322 -----GluIleAspHisMet--IleGluThrGlyProSerValGluAsp 335
 QY 2656 ATTGTTGATGATTGATTAATAATTCCTTTTGGGGGATGTTCTCTACGCAAAATAT 2715
 Db 336 IleSerGluAspLeuValLys-----LysTyr 344
 QY 2716 ---GAAGCAACAAAGATATACGATGTATAAGAA-----AAGATTAACCTAAAGGC 2766
 Db 345 SerGlnLysAsnAsnMetIleLeuLeuSerAsnAspSerTyrLysAsnLeuLeuLys 364
 QY 2767 CCCAAAGAAAGTACTGACCCCAACACAGACAACTATCTGTGAGCTATACGTATACGT 2826
 Db 365 SerGlnSerLysSerLysProLysAspArgLysLeuMetThrLysGluVal----- 381
 QY 2827 TTTCGATATATAGTATATATTCGAGAGAGATCTCTGGGAAAGAAACGGTGACATG 2886
 Db 382 -----AlaGluAsnLeuAsnMet 387
 QY 2887 GTAAGGTGCAAGACATTTGAAACTGTTTTGTAATATATACATAAAGTCACTCAAGGC 2946
 Db 388 IleAlaLeuPro----- 391
 QY 2947 AAAGCAATGATTAATATATGATGATGCCCCCAATATTTAAATGAGGCAAAATTTGG 3006

Db 392 -----AsnAspAsp----- 394
 QY 3007 TGGGAGCTAATATAGACCACAAAGATATGGAAACCATGAATGTGATTAATAATTTGAG 3066
 Db 395 -----AsnTyrSerLysLysGluPheSerLeuGlnSerHisIleLysTyrLeu--- 410
 QY 3067 GATTAATCGGACACACCA-----TCACACAAAGATGATTTGGCGATATATGAT 3117
 Db 411 GluAlaSerGlyTyrLysValLeuProLeuGluGluPheGlnAsnLeuAsnGluSerLeu 430
 QY 3118 CATACCCCTTGATGATATATATCCACAAAATTAAGATGATGACCAATGGCGACGA 3177
 Db 431 SerAsnProSerTyrAsnTyrLeuLysGluLysLeuGlnAlaLeuLysLys----- 447
 QY 3178 TGGTACTGCAAGGTGACGAAAAAGAGTGTGATTAAGTGAACAG----- 3222
 Db 448 -----IleProIleAspIleSerThrPheAsnLeuLeuLysGluProThrIleAspPhe 465
 QY 3223 -----AAGTGAAGAGTGAATGAGATATAGATATAGT--- 3255
 Db 466 LeuLeuProLeuThrSerLysIleAspCysLeuIleIleProThrLysAspTyrAsnAsp 485
 QY 3256 -----CAAGCTGTACGAAAGAGAGTGTACAGCTGTACGAACTGCAAGAGCTGT 3309
 Db 486 LeuPheGlnSerValLysAsnProSerIleGlnGlnMetLysLysCysLeuGlnAlaLys 505
 QY 3310 AATGAATATATATATATATATGATTAATGGAAGAACATGCAAT----- 3354
 Db 506 AsnAspLeuGlnSerAsnIleCysLysThrLeuGlnGluArgAsnGlyCysLysThrLeu 525
 QY 3355 -----ATAATA 3360
 Db 526 SerAsnAspLeuTyrPheSerMetValAsnLysIleGluThrProSerLysGlnTyrLeu 545
 QY 3361 TCAGTAAATACAAAGATTTACATGACACACAAATGTCTGATGATATAGTGATTT 3420
 Db 546 SerAspLysAlaLysGluTyrAsp-----GlnValLeuIleAspThrLysAlaLeu 562
 QY 3421 GAAAGCTTCAGTACTGCCAAATATCATATGACAGAGATGTTATGAAATTTTGTGGAA 3480
 Db 563 GluGly-----LeuLysAsnProThrIleAspPheLeuArgGln 575
 QY 3481 TTATACCAAAATGTCGCAAAAGATTAATAAGTGTACTAGTAT----- 3528
 Db 576 -----LysAlaSerAlaSerAspTyrLeuLeuLeu 585
 QY 3529 ---GAAAGTGTGCTGATTTGCTACTAACACACAGTAT-----GAAATGTGGAGCA--- 3576
 Db 586 LysLysGlnAspTyrValSerProSerLeuGlnIuThrLeuValGlnHisAlaLysAlaThr 605
 QY 3577 ---TATCTCATATGATACAGAAATTTTGATGAT-----TGTCAGTCA----- 3615
 Db 606 AsnHisHisLeuLeuSerAspSerAlaTyrGluAspLeuValLysCysLysGluAsnPro 625
 QY 3616 CAATATGATTTTGTGATGAAAAAGT-----GATGTGAAGATTAACGAA 3660
 Db 626 AspMetGlnPheLeuLysGlnLysSerAlaLysLeuGlnLysHisThrValSerAsnGln 645
 QY 3661 AAATATGCTTTTACAGATTAACACAGACCAATATGTCGCTGTGTTAAAGTGA 3720
 Db 646 AlaTyrSerGluLeuGlnLysLysLeuGlnGlnProSerLeuGlnIuThrLeuValGlnHis 665
 QY 3721 TCGAAACCGCACAGGATACAGATTAACGAAAAAAGCGAAGAAAGATACGGA 3780
 Db 666 AlaLysAlaThrAsnHisHisLeu--LeuSerAspSerAlaTyrGluAspLeuValLys 684
 QY 3781 TGTAAGACAGTGAATGATTA-----CTTAAGAAACATGCAAGAAACAGTAGAA 3834
 Db 685 CysLysGlnAsnProAspMetGlnPheLeuLysGlnLysSerAlaLysLeuGly----- 702
 QY 3835 GATTTGATCCAAAAAAGATAGATATGATATCCGATTCGGCATGCGGAATGATTAAT 3894
 Db 703 -----HisThrValIleSerAsnGlnAlaTyrSerGlnLeuGlnArgLysTyrSerGln 720

QY 3895 TTA-----GTGAAGACCCCTGTGTGTATGCCCCCTAGAACACAAAGTTATGC 3945
 Db 721 LeuGluValGluGluInProSerLeuAlaTyrLeuValGluHisAlaTyrSerLeu 740
 QY 3946 GTTCATTTCTTGGCAATGATTAATGAATTAATAAATTTACATACACAGTTAATTTAAA 4005
 Db 741 AspHisHisLeuLeuSerAspSerAlaTyrGluAspLeu-----ValLysCysLys 757
 QY 4006 GAA-----GCTTCATCAATCTGCAGCAGCAGAA-----ACATCTCTC 4044
 Db 758 GluAsnProAspValGluPheLeuLysGluLysSerAlaLysLeuGluHisThrValVal 777
 QY 4045 TCATGTAATTATTAATAAGTAAGATGTGAAGAAATGACATCGATAAAGATTAAAA 4104
 Db 778 SerSerGluGluTyrSerGluLeuGluArgLysTyrSerGluLeuGluLysGluValGlu 797
 QY 4105 GAAGGCAAAATTCCTCCGATTTTGCAGATCCATGTTCTACACATTTGGACAGATTAGA 4164
 Db 798 GluProSerLeu-----AlaTyrLeu-----Val 805
 QY 4165 GATTTTATTTATTTGACACAGAT-----ATATCAAAAGGTGATGTTGAGGAGAT 4212
 Db 806 GluHisAlaLysAlaThrAspHisHisLeuLeuSerAspSerAlaTyrGluGluLeuVal 825
 QY 4213 AAACATAAGACCAATAGATTCTCTTTTCT---AAAAATGTCACCAAAATCTCTAAT 4269
 Db 826 LysCysLysGluAsnProAspMetGluPheLeuLysGluLysSerAlaLysLeuGluHis 845
 QY 4270 GGAAGAAACAGCCAGAAATGTGGACAGACATAGTCATGATATTTGGAAAGCTATGCTA 4329
 Db 846 ThrValValSerAsnGluAlaTyrSerGluLeuGluLysLysLeuGluInProSerLeu 865
 QY 4330 TGTGCACTAGTAATAAATTTGGGCAAAAAAAGATGATTTTACGAAAACTACGGTTACAC 4389
 Db 866 AlaTyrLeuValGlu---HisAlaLysAlaThrAsp----- 876
 QY 4390 AACGTCAAAATTTAGTCACAAAGACACCATTGGAGAAATTTGCCAAACGACCCCACTTT 4449
 Db 877 HisHisLeuLeuSerAspSerAla----- 884
 QY 4450 TTACAGATGGCTAACCGAATGTGACAGACACTATTGCTATACAGACAAAAATTTTGAAG 4509
 Db 885 -----TyrGluAspLeuVal----- 889
 QY 4510 GATGTGAGCAAAAAATGTAGTCAAAATGACCAATTTGAAGTGTGATACAGAAATGATAAG 4569
 Db 890 -----LysCysLysGluAsnSerAsp----- 896
 QY 4570 AAATGCGAGACTAACGTTAATATATATGAAAAA-----GAGTGGATT 4617
 Db 897 -----ValGluPheLeuLysGluLysSerAlaLysLeuGluHisThrVal 911
 QY 4618 CCACAAATGAATATTTCAAGATGAACGCGCAAAAAAAGATTCGTAGACAAACATATC 4677
 Db 912 ValSerAsnGluAlaTyrSerGluLeu-----GluLysLysLeuGluInProSerLeu 929
 QY 4678 GGTGTATGTGTACAGACTACTGGAACCAATGACCAACAGATTACTGAAACAGAAATTT 4737
 Db 930 AlaTyrLeuValGluHis-----AlaLysAlaThrAspHisHisLeuLeuSerAsp 946
 QY 4738 ACTGCT-----ACTGTGTGTATGAAGCTTGGAAAGTGGCTGTGTGTACAA 4782
 Db 947 SerAlaTyrGluAspLeuValLysCysLysGluAsnProAspMetGluPheLeuLysGlu 966
 QY 4783 AGCAATATACAAATTTG-----TTAGAAAAACAGCCTTATGATGGCCGCAAAA 4830
 Db 967 LysSerAlaLysLeuGluHisThrValValSerAsnGluAlaTyrSerGluLeuGluLys 986
 QY 4831 -----CATTTGGGTGTCACAAAA---TTT 4851
 Db 987 LysLeuGluGluInProSerLeuGluTyrLeuValGluHisAlaLysAlaThrAspHisHis 1006

QY 4852 ATTGAATATGACGACAAATATTAACATTTTGCAGTAAGAAATGATGCAAAAGA----- 4905
 Db 1007 LeuLeuSerAspSerAlaTyrGluAspLeuVal-----LysCysLysGluAsnPro 1023
 QY 4906 -----TTAGTAAGGAGGCAAAACAGAGTCTCTTTAAGTGGCAAAACAA 4950
 Db 1024 AspMetGluPheLeuLysGluLysSerAlaLysLeuGluHisThrValValSerAsnGlu 1043
 QY 4951 GGTCCTAATATCAATTAACATACTGAAAGATGACAGATGAGCTTTTCTCTCT--- 5007
 Db 1044 Ala-----TyrSerGluLeuGluLysLysLeuGluIn-----ProSerLeu 1057
 QY 5008 -----CGTCAGCTACGATATATTTTCATGCAATG---GATGGCAATAT 5049
 Db 1058 GluTyrLeuValGluHisAlaLysAlaThrAspHisHisLeuLeuSerAspSerAlaTyr 1077
 QY 5050 ACAGATCCACAAATTAAGATGAAGAAATGGTTCGCAAAAAAGATGATGAAAGGCGCGCA 5109
 Db 1078 GluGluLeuValLysCysLysGluAsnProAspValGluPheLeuLysGluLysSerAla 1097
 QY 5110 ACGGAAGGTCACAAATTTG-----GGTCAATACTACAAAGAAAAAAGAAAAA----- 5157
 Db 1098 LysLeuGluHisThrValValSerAsnGluAlaTyrSerGluLeuGluLysLysLeuGlu 1117
 QY 5158 -----GAGAAATTAATAACCTGCGATGCGCCACAAATATTTCT 5193
 Db 1118 GluProSerLeuGluTyrLeuValGluHisAlaLysAlaThrAspHisHisLeuLeuSer 1137
 QY 5194 TATGAGGTCCCGCTTGTATGCTATGTAATATAGTTTATATGATTAAAGATATATAT 5253
 Db 1137 ----- 1137
 QY 5254 CTAGTATGATATATTTGGAAGATGAAGAAACAAAGACGAGAAATTTGAAGAAATA 5313
 Db 1138 -----AspSerAlaTyrGluGluLeuValLysCysLysGluAsnProAspValGlu 1154
 QY 5314 TTTTAACAAAAATGGAACATGATGTTGCGCAAGAGATGATGTCTACAGCAAAATCCCGGT 5373
 Db 1155 PheLeuLysGlu---LysSerAlaLysLeuGluHisThrValValSerAsnGluAlaTyr 1173
 QY 5374 AGTACTGCGGGAATTTTCTGGACGAAATTAAGAAATGTGTGTGGAACGCAATGATA 5433
 Db 1174 SerGluLeuGluLys-----LysLeuGluInProSerLeuAlaTyrLeu 1188
 QY 5434 TGGGGGTACAAAGCTGTAGGATGATGAAATAGTGAATATGTCGCAAGAGTGAATG 5493
 Db 1189 ValGluHisAlaLysAlaThrAspHisHisLeuLeuSerAspSerAlaTyr-----Glu 1206
 QY 5494 GATCTAAAAAATGTGCTGTGTAACCTTCAGATGATGATATTCCTATGGGAAAAATCGC 5553
 Db 1207 AspLeuValLysCysLysGluAsnProAsp----- 1216
 QY 5554 GATGAAGTACTCGGTCAAGTTCTTGATGGTTTCCGAAATGGGCTGAAGATTTTGC 5613
 Db 1217 -----ValGluPheLeu 1220
 QY 5614 AAACATTAAGAAAAAGATTTGAGAAATTTGTAGGGCTGTATGATTACTTGTGT 5673
 Db 1221 LysGluLysSerAlaLysLeu-----GluHisThrValVal 1232
 QY 5674 GATTAATGACATTAAGAAAGAAATTTACAGATGCGGTGTACACATATTAATAATTTAT 5733
 Db 1233 SerAsnGlu----- 1235
 QY 5734 AGTGAAGTGAAACACAGATGAAGAAACAAATCAAAATATGCTGACAAATTAAGCAAA 5793
 Db 1236 -----AlaTyrSerGluLeuGluLysLys 1243
 QY 5794 ATATATTCGAGCATCCTGTGGCAAAAAGATGCAAGAGACGCTCGCAATTTTATGACAAA 5853
 Db 1244 Leu-----GluGluInPro-----SerLeuAlaTyrLeuValGlu 1254
 QY 5854 CAATTAATAAAATTTTGTGAATAATAAGATGAGATTTGTGAATAT-----AAG 5901

```
Db 1255 HisAlaLysAlaThrAspHisHisLeuLeuSerAspSerAlaTyrGluAspLeuValLys 1274
QY 5902 TGTATGAAGAATGTGTCCACACACAGCATTAAGTATGATATAGTCAAAATATGCCCGCA 5961
Db 1275 CysLysGluAsnProAspMetGluPheLeu----- 1284
QY 5962 TCATTACGATGACCAAAAGAAAGTTGAAAGAAAGTGAATGTCAAGTCCACGAGGT 6021
Db 1285 -----LysGluLysSerAlaLysLeuLysHisThrVal----- 1295
QY 6022 CCAACGCTGACGAGGAGAAACACCGTCAACGCGGTATCAGTATTCAGAAAGCAGC 6081
Db 1296 -----ValSerAsnGluAla 1300
QY 6082 GCATCGAAAAAGAGCGAAACAGCCCGCTACAAACAGCCGAAAAAGTGAATAAT 6141
Db 1301 TysSerGluLeuGluLysLysLeuGluProSer-----LeuGluTyr 1315
QY 6142 CTAAACAAGAAATGGAGGACAAACAGAACCCGACGACACACAAACAGAA 6201
Db 1316 LeuValGluHisAlaLysAla----- 1322
QY 6202 CGAACATCAACAGCAACACACAGATCTAGCTGGGACATGGTAAAGCCATTCTT 6261
Db 1323 ---ThrasnHisLeuLeuSerAspSerAlaTyrGluAspLeuValLysCys----- 1339
QY 6262 TCGAATTAACACATACGACGAGGCTGCAATGAGCGGTGTATCCAAACGATATGACAA 6321
Db 1340 LysGluAsnProAsp-----MetGluPheLeuLysGluLysSer----- 1352
QY 6322 TATCTAAATGGGGTGTATGTAGTAAAGTCAAAAGAAATGCAATATGTATG 6381
Db 1353 ---AlaLysLeuGluLysHisThrValValSerAsnLysGluTyrSerGlu----- 1367
QY 6382 CCTCCTAGAGAAAAAATTATGTATTAATATATACAAATTTAAATTAATGAAACTGA 6441
Db 1368 ---LeuGluLysLysLeuGluGlnProSerLeuGluTyrLeu----- 1380
QY 6442 AATAACCGTACACATGATTAAGAGAGCTTTATTAATATGTCGACCAATACAACTCA 6501
Db 1381 ValLysHisAlaGluGlnLeuGlnSerLysLysLeuSerLysSerAspPheAsnThr--- 1399
QY 6502 TTTTGTGGTTAAATATATTAATTAAGAAATCTGCAGCAAAATGAATGCAAAATGA 6561
Db 1400 -----LeuAlaAsnProSerMetGluAsp----- 1407
QY 6562 ACAATTCAGATGAATTAAGAAATATATGTATATCATATGTGATTATTAAGATATG 6621
Db 1408 ---MetAlaSerLysLeuGlnLysLeuGluTyrGlnIle----- 1419
QY 6622 TTTTGTGAACATGATATTTCTAATATAAATAAATTAATA-----ACGTAAACAAT 6672
Db 1420 -----ValSerAsnAspGluTyrLysLeuLysHisThrMetGluLys 1434
QY 6673 AGTGAACACCATCTCAATGAAAAATTAATAAGAAAAAAG-----GATAAAAA 6726
Db 1435 ProAspValGluLeuLeuAspSerLysLysGluTyrHisLysLysLeuAspThrThr 1454
QY 6727 GATGAAGATTAAGTAAATATTTGGAGAAATAAATAAATTTATTTGGAGAAATG 6786
Db 1455 TyrAsnGluLeuValSerAsnPheAsnSerProThrLeuLysPheIle---GluGluLys 1473
QY 6787 ATATATGATTAACCTATCATCTCACAGACGAAACGAAAAAAGAAATTAAGATATAT 6846
Db 1474 AlaLysSerLysGluTyrArgLeuLeuGluProAsnGlu----- 1486
QY 6847 TACAGTTCATGACATGACAAACTG-----ACGCTTCCTGAGAGTTGTAAAA 6900
Db 1487 -----TyrLeuAspLeuAsnArgLysLeuAlaThrThrProSerLysGluGlnIle----- 1502
QY 6901 AGGCCCAATTTTGAATGTGTACAGAAAGGCGAGAAAGATTTTGAATTAAGAGAAAG 6960
Db ----- 1511

Db 1503 -----AspAsnPheCysLys----- 1507
QY 6961 GAACAGTTGTTAAATTTGAGCGCGGCTGT-----AAGCAATATGAG--- 7002
Db 1508 -----GlnIleGlyCysTyrAlaLeuAspSerLysGluTyrGluArg 1521
QY 7003 TGTATGTAGTATGACGCGTAAACAGAAATGTCGACGAGCGGTCTTAACATATCA 7062
Db 1522 LeuLysAsnSerLeuGluAsnProSerLysLysPheIleGluGluAsnAlaLeuLeu 1541
QY 7063 AATTTATTAAAGAGTGAAGAACTGAATATGAAGACAAAGAAAGTTCAAAAAAGAT 7122
Db 1542 AspLeuValLeuValAspLysThrGluTyrGlnAlaMet----- 1554
QY 7123 AAAGATGGCAAAAAGTATAGATATATCTCTACTGAAGACATGACAGCAACA 7182
Db 1555 -----LysAspAsnAlaSerAsnLysSerLeuIleProSerThr 1568
QY 7183 TGTGCTCATGATATTTAAACATGAATTAAGAAATTAATGTCGCAATAGGATGTCT 7242
Db 1569 LysValLeuAspPheValThrMetProAlaProGlnLeuAlaSerAlaGlu----- 1585
QY 7243 TGTATGCAAAAACCTTTCTTACACACTACCAAAACACAAACAAATCATCCGAT 7302
Db 1586 -----LysSerSerLeuGluLysArgThr-----LeuSerAsp 1596
QY 7303 GCTAATGATATGCGCAAAATCGCTGATATGTCTCTGAAGAAATTAACAGTGCAGGT 7362
Db 1597 IleGluAsnGluLeuLysAlaLeuGluTyrVal-----AlaIleArgLysGluAsnLeu 1614
QY 7363 CCTGAACCTTCAAAAAGGAGATCTATGATTCATACAAAAAATTAAGTCAATAATA 7422
Db 1615 ProAsnLeuGluLysPro----- 1620
QY 7423 CCTATGAATGTGTAGAGAAAGCAGCATATTTATCTAAAGACGCAAAATATATG 7482
Db 1621 -----IleValAspAsnAlaSerLysAsnAsp 1629
QY 7483 GATATTAACCTTGAGAAAAATTT-----ATACCTATGAGTCTACAAAGCAAAAGAA 7536
Db 1630 ValLeuAsnLeuCysSerLysPheSerLeuValProLeuSerThrGluGluTyrAspAsn 1649
QY 7537 AGTAAAAATATGTGACATAT-----ATATACTCTGCGATCCTTAAGAAACCTTATGA 7590
Db 1650 MetArgLysGluHisThrLysLysLeuAsnLysLeuGluLysProSerLysAspPheLeu 1669
QY 7591 CCGTAAATATATATAGAGAAAGAACCCCTGTGAAAAATAGAGAAAGAAATCGTTTATG 7650
Db 1670 LysGluLys-----CysGlu----- 1674
QY 7651 GTACATTTATGATGAAATGTTTACAAAAATTCAAAGTTC---TATCAGAGAAAAAAGA 7707
Db 1675 -----LysTyrGlnMetLeuIleIleSerLysHisAspTyrGluGluLysGlnGlu 1691
QY 7708 GTATGTACCTCCAGAGAGACATATAGCTTAAGSAT----- 7749
Db 1692 AlaIleGluAsnProGluTyrGluPheIleLeuGluLysAlaSerAlaLeuGluTyrGlu 1711
QY 7750 ---TTAATGAAATTAATAATGGAAGACTTAAG-----GATAGTAATTAATTCCTTA 7797
Db 1712 LeuValSerGluValGluLeuAspArgMetLysGlnMetIleLeuAspSerProAspIleAsp 1731
QY 7798 AAAATGTTGCTGCAACTGCACGCAAAATGAAGAAATGACATTAATTAATAAC----- 7848
Db 1732 TyrMetGlnLysAlaAlaArgAsnGluMetVal---LeuLeuArgAsnGluGluLys 1750
QY 7849 -----TTCAATCAGAGAAAGGCTGC 7869
Db 1751 GluAlaLeuGlnLysLysLysGluTyrProSerLeuThrPheLeuIleGluLysAlaAla 1770
QY 7870 GCAATGAATCCAATA---TGTGATACTAGTAATATAGTTTCCTGATC----- 7915
Db 1771 GlyMetAsnLysLysLeuValAspGlnIleGluTyrAsp---GluThrIleArgLysCysAs 1790
```

QY 7916 -GGGTGACACTACTGTAGAC---GAACAGATATGTACGAATTCGGTGTACTTACCTCC 7971
DB 1790 nHISProthrArgMetGluLeuGluSerCysHISHisLeuValLeuLeu-- 1809
QY 7972 GTGCAATATAATATATAGCTTTTGTATACATATATATGGAAGAAATATAAAT 8031
DB 1810 ----AspGlnAsnGluTyrSerThrLeuArgGluProLeuLysAsnArgAsnValGlnAs 1828
QY 8032 AAGGTAGAAATATATACACAGATGTACAAACGTTTCGTTGCTGCTGGAGTGTAT 8091
DB 1828 pleu-----LleasnThrLeuSerLysLeuAsnTyrLleAlaLleProAsnThrLleTy 1846
QY 8092 AGAAAAATATTTTGAAGCAATGACGTGCAAAACACACAGACATGCAAACTTTTACA 8151
DB 1846 rGlnAspLeuLleGlyLysTyrGluAsnProAsnPheAspTyrLeuLysAspSerLeuAs 1866
QY 8152 AAGGAAGAATGATGATTTTGAACCATATACATTAATTAACAGATAGCTG----- 8203
DB 1866 nLysMetAspTyrValAlaLleSerArg--GlnAspTyrGluLeuMetValAlaLysTy 1885
QY 8204 -----GACATTAAGACGATC-----CACCTGTGTAT 8229
DB 1885 rGluLysProGlnLeuAspTyrLeuLysLleSerSerGluLysLleAspHisLleValVa 1905
QY 8230 GATTATATACCTCAGCGTTTCGATGGAATGCTGTAATATTTATTTAAGCA 8289
DB 1905 lPro-----LeuSerGluTyrAsnLeuMetVal----- 1914
QY 8290 CTGATGGAAGATTCGAAAAAT-----TTAA 8316
DB 1915 -----ThrAsnTyrArgAsnProSerLeuSerTyrLeuLysGluLysAlaValLeuAs 1932
QY 8317 AATATCATGTATC----- 8329
DB 1932 nAsnHisLleLeuLleLysGluAspAspTyrLysAsnLleLeuAlaValSerGluHisPr 1952
QY 8330 -ACTGTAACATCTGACAGATGCAAGATGATTATGANGAATAAGTGTAC----- 8383
DB 1952 oThrValLleHisLeuSerGluArgHisLeuLeuAsnLysValLeuValAspArgAs 1972
QY 8384 -----AGTGTAAACGAGATGCAAGATATATAAATTTGCTTAATAGS--- 8430
DB 1972 nAspPheAlaThrMetSerArgSerLleGluLysProThrLleAspPheLeuSerThrLy 1992
QY 8430 ----- 8430
DB 1992 sAlaLeuSerMetGlyLysLleLeuValAsnGluSerThrHisLysArgAsnGluLysLe 2012
QY 8431 -----AAATCTGTATTCGA 8444
DB 2012 uLeuSerGluProAspSerGluPheLeuThrMetLysAlaLysGluGlnGluLleLle 2032
QY 8445 TATACATCAATTAATATACAAAGATTTGTATGAACA-----CC 8483
DB 2032 eLleSerGluLysGluTyrSerGluLeuArgAspGlnLleAspArgProAsnLeuAspVa 2052
QY 8484 AATATATACAAATCTCTACTATGATCAT-----GTCAAAATTTT---GTACAAA 8534
DB 2052 lLeuLysGluLysAlaLlePheAspSerLleLleValGluAsnLleLleLysGlnG 2072
QY 8535 GTTGAAAACTTTTAAAGCAATGT-----TCTGTTGAGACGTTTTCGATATCTC 8585
DB 2072 nLeuValAsnThrThrSerProCysProLleThrTyGluAspLeuLysValTyAl 2092
QY 8586 TCATGAA-----ACAAGTAAAGTTTGAATTAATTAATTAAGAAATGATGTTCTTC 8639
DB 2092 aHisGlnPheGlyMetGluLeuCysLeuGln---LysProAsnLysLeuSerGlyAla-- 2110
QY 8640 CAATATACGAACATATGCTTCGAGAAACCAAAAGTTATTAAGAACGCTTCAGTTG 8699
DB 2111 -----GluArgAlaGluArgLleAspGluGlnSerLleAsnThr 2123

QY 8700 TACACTACCTTCTAAGATTCATGGATTAATTTCTACCGATCAAAACAAAGATGATG 8759
DB 2123 rThrSerSerAsnSerThrThrThrSerSerMetPheThrAspAlaLeuAspAsnL 2143
QY 8760 TAAGGAATTCAAACCTTTTACCTTCCTGCGAAGAAATGATTATGATTAATTAATTCGTA 8819
DB 2143 eGluGluLeuAsnArgValGluLeuGlnAsnAsnGluAspTyrThrAspLleIleSerLy 2163
QY 8820 TTGGAAC-----GCATACCTTGTCTTAATATTC 8849
DB 2163 sSerSerThrValLysAspAlaThrLlePheLleProLalaTyGluAsnLleLysAsnSe 2183
QY 8850 AGATGATTAACAAAGCTGTA---TTGATTCCTCCAGACAGACATTTATGTCAAGACC 8906
DB 2183 rAlaGluLysLeuGlyTyrLysLeuValProPheGluLysSerAsnLleAsnLeuLysAs 2203
QY 8907 TATCAGCTCATATTAATATGAAAGCGTATTAAGAAATTTTAAAAAAACTTCTTAC 8966
DB 2203 nLleGluAlaProLeuPheSerLysAspAsnAspAsp-----Th 2216
QY 8967 TTCTGCTTCAGTCAAGACAAATGTTAGTCAAAATATATATTCGAGACAGATGTTG 9026
DB 2216 rSerValAlaSerSerLleAspLeu-----AspHisLeuSe 2228
QY 9027 CTTTGAGCGCAATGATATATGTTATGCACATTTTCCGAT----- 9066
DB 2228 rArgLysAlaGluLysTyrGlyMetThrLeuLleSerAspGlnGluPheGluGluTyrH 2248
QY 9067 -ATATATTA-----GGACGATATATGACACTTCAT 9101
DB 2248 sLleLeuLysAspAsnAlaValAsnLeuAsnGlyGlyMetGluGluMetAsnProLe 2268
QY 9102 ATCTGAAAAATTAATAAATATTTGAAACATCAATATGAACACGAAATGTAAMAC 9161
DB 2268 uSerGluAsnGlnAsnLeuAlaLysThrThrAsnThrAlaGlnGluLy----- 2285
QY 9162 ATGCTGGAAAAATATATAGACGTACATATGCGACCTATGTATGCGATATTAATTC 9221
DB 2286 -----AlaPheGlnAsnTh 2290
QY 9222 TACTCAAAATGTAATATGATGAGAGATGTCATATCCAAAGATGAAACCTAA 9281
DB 2290 rValProHisAsnAspMetAspAsnGluGluValGlyTyrGlyProAspAspProThr-- 2309
QY 9282 TCAGTTTCTTCGTTGTTAATGAAATGCGCAAGACAGATGTAAAGAAACATGT 9341
DB 2310 -----PheThrValArgGlnLeuLysLysProAlaGlyAspArgAsnLeuL 2325
QY 9342 AAGTGATTCATTAAAAACAATAATGTCCTGTCAAACGAAATATTT----- 9390
DB 2325 eLeuThrSer---ArgGluLysThrLeuLeuSerArgAspAsnLleMetSerGlnAs 2344
QY 9391 -----GAAGCGTCGAGATTTAT 9407
DB 2344 nGluAlaValTyGlyAspAspLleSerAspSerPheValAspGluSerGlnGluLe-- 2363
QY 9408 AAGCAACCTGATGTCAGATGATAT-----AGAAATATAT 9446
DB 2364 -----LysAsnAspValAspLleLleLysThrGlnAlaMetLysTyG 2378
QY 9447 TACCTTGAAATATATGTAATAAATATCAATGAAATCTAAATATATAATTAAGCAAT 9506
DB 2378 yMetLleCysLleProGluSerAsnPe-----ValGlyAlaSerTyrAlaSerAl 2395
QY 9507 AAAAGATCAATCTCA-----GATAATATAGAC----- 9534
DB 2395 aGlnAspMetSerAspLleValValLeuSerAlaSerTyrTyHisAsnLeuMetSerPr 2415
QY 9535 -----ATAAACCATCTGAAGAAATGTTCAAGTCAATTAATAATCAAA 9578
DB 2415 oGluAspMetLysTyrAsnCysValSerAsnGluGlnLeuGlnAlaGluValLysLysAr 2435
QY 9579 AGATTCTCAATGCGCTTGGAGTTAAATGAT-----AT 9611

```

Db 2435 GGLYLeuGlnIleAlaLeuThrThrLysGluAspLysGlyGlnAlaThrAlaSerLys 2455
OY 9612 AAATGAATAGTATGACGAGCAAAAAAATATGAATAATGAATATC----- 9657
Db 2455 SHISGLuTYValSerHisLysLeuAanHisLysThrSerThrValSerThrLysSerG1 2475
OY 9658 -----AAGAAGTACTAAAAAATATATATCCGTGTTATATTTGTTGAGATGAAACACA 9713
Db 2475 YAlaLysLysGlyLeuAlaGlnAlaAlaThrThrAlaLysGluAspSerGlnSerH1 2495
OY 9714 CAAAAATCAGTACTAGATGCAATATATATAAAGAAAGACCAAAAGCTGCTCTTAAGC 9773
Db 2495 SPro-----GlnIleGluLysLysSerHisArg----- 2504
OY 9774 ACTCTATTCTTACACCCCATGATGATTTCTTATCAACACACCTTATTTCAACACA 9833
Db 2505 -----ThrAsnHis-----HisLysHisH1 2511
OY 9834 TCGAGTAGCACATATGATCTCTAAAAATGATATATGAAAGTAGATCTCTGTT----- 9888
Db 2511 sLysArgGlnGlnSerLeuAanSerHisSerThrSerLysThrHisSerSerArgAs 2531
OY 9889 -GTTATTTGATTCGCGCTGTTAGGTTGATGAGCGCTTCATTTTCATGACAGAA----- 9936
Db 2531 nThrProAlaSerArgArgAspIleValAla-----SerPheMetSerArgAlaGlySerAl 2550
OY 9937 -AAATTCAAATCGCTGTGGACTGTTGGCTATGATGAAATATCCGCA----- 9984
Db 2550 aSerArgThrAlaSerLeuGlnThrLeuAlaSerLeuAanGluProSerIleIleProAl 2570
OY 9985 -----GGAGATGATGAAATGCTAGCTAGTGGATGCAAAATATGCTA 10025
Db 2570 aLeuThrGlnThrValIleGlyGlnLys-----LeuPheLysThrLysProAlaGlyGln 2589
OY 10026 CATACCATATAGAGTGTCCATATATAAGCAAAACATATATATATATG 10075
Db 2589 rObheGlyPheGlnSerArgHisGluArg-----PhePheThr 2601

RESULT 13
MST2_DROHY
ID MST2_DROHY STANDARD: PRT; 1391 AA.
AC 008696;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Axoneme-associated protein mscl01(2).
GN MScl01(2).
Drosophila hydei (Fruit fly).
Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RA MEDLINE=95045538; PubMed=7957199;
RA Neesen J., Padmanabhan S., Buenemann H.;
RT "Tandemly arranged repeats of a novel highly charged 16-amino-acid
RT motif representing the major component of the sperm-tail-specific
RT axoneme-associated protein family Dmscl01 form extended
RT alpha-helical rods within the extremely elongated spermatozoa of
RT Drosophila hydei."
RL Eur. J. Biochem. 225:1089-1095(1994).
CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCTYTES AND EARLY
CC SPERMATIDS
CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
CC -1- POLYMORPHISMS: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
CC REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

DR EMBL: X73481; CAA51876.1; -
DR PIR: S34154; S34154.
DR Flybase: FBgn020733; Dhyd\mscl01(2).
KW Sperm; Repeat; Multigene family; Polymorphism.
FT 59 x 16 AA APPROXIMATE TANDEM REPEATS OF
FT DOMAIN 332 1268 [KR]-K-X-C-X-X-K-X-K-X-K-X-X-E.
SQ SEQUENCE 1391 AA; 159000 MW; 1B2A366F30E4878 CRC64;

Alignment Scores:

Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
4.23e-08	1391	344	230	631	554	80
Score:	319.50					
Percent Similarity:	32.63%					
Best Local Similarity:	19.56%					
Query Match:	1.64%					

US-10-087-013-1 (1-10628) x MST2_DROHY (1-1391)

```

OY 4519 GAAAAATGTAAGTCAAAATGACCAATGTAAGTGTGATACAGATGTAATGAAATGCGAG 4578
Db 50 GlnSerThrAspAspSerAsnGlnIleArgCysAspAlaAsp-----LysThrAlaAla 67
OY 4579 GACTACGTTAAATATATATGTAAGAAAAAAGAGATGATTCACAAAGATATATATACAG 4638
Db 68 Asp-----LysLysLysGlnLysGln-----Lys 75
OY 4639 GATGACCGCAGCAAAAAAAGATGCGATGACACACATGTTGTAATGCTTACAGACTAT 4698
Db 76 GlnGlu-GlnGlnAlaLysLys-----ArgLys 85
OY 4699 ACTGGAACGAATGC-----AACAGATTAAGTGAACAGAAATTAAGTCAAGT 4746
Db 85 YLysArgGlnCysLeuLysValGlnLysArgValIleAlaGlnIleIleArgCys-Ser 104
OY 4747 TGTGCTGATTAAGCCTCGAAGTGCCTGTGTGTACAAAGATATATATGTAAGAAAA 4806
Db 105 -GlyGlnLys-----AspArgIleLeuIleGlnLysMetLysCysLeuThr--- 120
OY 4807 CAGCTTACTATGATGCGCGACCAATGTTGGCGCAAAATTTATGAAATGAC--- 4863
Db 121 -----AspGlyMetLysLys---AlaCysThrIleAlaLysAlaLysLeu 135
OY 4864 -GACAAA-----TATACTAACATTTGAGTAAGATTAAGTGCAAA 4902
Db 136 IleAlaAspLysGlnLeuAlaValGlnCysAlaLalaLeuSerLysAspLysValLys 135
OY 4903 GGATTACTA-----AAGAGCGCAACACAGGTGCT 4932
Db 156 AlalaLeuLysLysCysGlnArgLysSerLysGlnLysCysAsnGlnAsnSer 175
OY 4933 -----ATTAAGTGGCAAAACAAAGAGTCTCTATTAATAC 4965
Db 176 ProAlaGlnLysAspLysAspArgThrLysLysGlyLysThrLysGlySerGlyGly 195
OY 4966 AATTAACCTGAAAGAAATGACTGAAGAT-----GTGCTTTTCT 5004
Db 196 GlyAsnLysLysArgSerThrLysGluAsnArgAlaLysLysGlyLysLysLeuValLys 215
OY 5005 TCTGCTGACAGCTATATGTTTTCATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 5043
Db 216 AsnArgPheThrGlnLysLeuGlnLysCysIleLysSerGlnLysAlaAspValCysGln 235
OY 5044 -AATTATACAGATCCAGAA----- 5061
Db 236 CysArgGlnAsnPheThrGlnAspGlnArgLysArgLysAlaLalaSerLysCysMet 255

```


Db 599 AspSerIuInLeuGluAsnIleLysMetAspLeuSerIleSerLeuGluSerIleGlu 618
QY 2248 AAACCTGGAGAGACGACCACTAAATAAATAAAGAAATCTTAA 2307
Db 619 AspProIuSerIleMetLysGluInThrLeuPheAspAlaGluThrValAlaLeuAspAlaLys 638
QY 2308 AGC---AGTCATACGAGAGACGAAATCGTCTTCATTAATTAAGAAAGGCA 2364
Db 639 ArgGluSerIlePheLeuArgSerGluAsnLeuGluLeu---LysGluLysMetLysGlu 657
QY 2365 CACGAGGTATATTAACCTGGGGGAGAGAAAGACTTCAAGCAATTAATGAGA 2424
Db 658 LeuAlaThrThrTyrLys----- 663
QY 2425 ATAATGATAAATCTTAATCTGAATCTGTTTCAATGACCATGATGCAAA 2484
Db 664 -----GlnMetGluAsnAsp-IleGluLeuTyrGln---SerGluLeuAlaLys 679
QY 2485 GCGACAGGTGATGTATACAAACACATTTGCTGAGAACGTAATGGAAAGTCGCG 2544
Db 679 sLysLysMetGluVal-----AspLeuGluLysGluLeuGluSerAlaPhe----- 694
QY 2545 GAACACATGCGTAAAGATACGAGAAATGTATATGCTCTGATAGAGACAT-ATATG 2603
Db 695 -AsnGluIleThrLeuThrSerLeuIleAspGlyLysValProLysAspLeuLys 714
QY 2604 TACATCTCAATTTGGAACATTTACAAACGAT---GATCACCCACTTAATGTAATGT 2660
Db 714 sAsnLeuGluLeuGluGluLysIleThrAspLeuGluLysGluLeuAsnLysGluValAl 734
QY 2661 TGATGATTTAGTAAATATCTTTGGGAGTGTCTTCATGACGAAATATGAAAC 2720
Db 734 uGlu-----AsnGluAlaLeuArgGluGluValIleLeuLeuSerGluLeuLys 751
QY 2721 AAACAGATATACGAATGTATTAAGAAAGATTAACCTTAAGGCCCCCAAGAAATAC 2780
Db 751 r-----LeuProSerGluValGluArgLeuArg-----LysGluIleGlu 764
QY 2781 TGAACCAAAACACGACACATCTGTGACATACCTTAACGAT---TTTGCAATAT 2837
Db 764 nAspLysSerGluGluLeuIleAsnIleThrSerGluLysAspLysLeuPheSerGluVal 784
QY 2838 A-----GGTGATATTAATTCGAGG-----AGAGCTCTCGGAAAGAAAGCGTGA 2882
Db 784 IValHisLysGluSerArgValGlnGluLeuLeuGluIleGlyLysThrLysAspAs 804
QY 2883 CATGCTAAAGCTGCAAGACATTTGGAACCTGT-----TTTGCTAA 2924
Db 804 PheuAlaThrThrGlnSerAsnTyrLysSerThrAspGlnGluPheGluAsnPheLysTh 824
QY 2925 TATACATTAAGTCACTCAAGGCAAA-----GGAATGATTAATATA 2966
Db 824 rLeuHisMetAspPheGluGlnLysTyrLysMetValLeuGluGluAsnGluLysGluMetAs 844
QY 2967 TGAATGATCCCCCAAAATTTTAAATGAGGAAATTTGTGGGAAGCTATAGACCCA 3026
Db 844 nGlnGluIle-----ValAsnLeuSerLys 852
QY 3027 AGTATGGAGCAAGCATGATGATATAAATATTGGAAGATAATCGGACACCAATC 3086
Db 852 s-----GluAlaGlnLysPheAsp-----SerSerLeuGluAlaLeuLys 865
QY 3087 AACACAAAGTATGATGCGGATATAGTATACACCATTTGATGATTAATCCCA 3146
Db 865 sThrGluLeuSer-----TyrLysThrGlu 873
QY 3147 AAAATTAAGATGATGACCCGATGGCAGAAATGCTACTGCAAGTGCAGAAAG----- 3201
Db 873 nGluLeuGlnLysThrArg-----GluValGlnGluLysGluLys 887
QY 3202 -GAGATGATAGTTGAAGGAGAGAGTGAAGAGATGTAAGATTAAGTGTCAAGG 3260
Db 111 ----- 111

Db 887 nGluMetGluGlnLeuLysGlu-----GlnLeuGluAsnArgAspSerProLeuGlu 904
QY 3261 CTGTACGAAGAAGAGTGTACAGGTTGTACAGATGACACAGAGCTTGATTAATATA 3320
Db 904 nThrValGluArgGluLysThrLeuIleThrGluLysLeuGlnGlnThrLeuGlu----- 922
QY 3321 TGAATTAATAGATTTGGAAGCAATGCAATTAATATACATTAATTAACAAAGATT 3380
Db 923 -GluValLysThrLeuThrGlnGluLys-----AspAspLeuLysGluLe 937
QY 3381 ACATGAACAGACCAAAATGCTGTAGTAATAGTGAATGAGCTTCAGATGCCAA 3440
Db 937 uGlnGluSerLeuGlnIleGluArgAspGlnLeuLysSerAspIleHisAspThrValAs 957
QY 3441 AAATCATATAGC-----AGCAATGTAATGAAATTTTGTGGAATTAATA 3485
Db 957 nMetAsnIleAspThrGlnGlnGluLeuArgAsnAlaLeuGluSerLeuLysGln----- 975
QY 3486 CCAACAAATGCTGCAAAAGTAATAAAGTGTACTGTAGTGAAGTGCCTGCATTCG 3545
Db 976 -HisGlnGluThrIleAsnThrLeuLysSerLysIleSerGlu----- 990
QY 3546 TACTAACACACGATGAAAATGTTGAGCATATCTCAT-----GATACAGGAAA 3596
Db 991 -----ValSerArgAsnLeuHisMetGluGluAsnThrGlyL 1003
QY 3597 TTTT---GATGATGTCAGTCAAAATGAGTTTGTGATGAAA----- 3639
Db 1003 uThrLysAspGluPheGlnGlnLysMetValGlyLleAspLysLysGlnAspLeuGluAl 1023
QY 3640 -----AGTATGTAAGATTAACGAAATATGCTTATAGAGA 3677
Db 1023 aLysAsnThrGlnThrLeuThrAlaAspValLysAspAsnGluIleLeuGlnGlnArg 1043
QY 3678 TAA-----CCACAGACATGATGCGTGTGTTGTAAGTGAATCGAA 3725
Db 1043 gLysIlePheSerLeuIleGlnLysAsnGluLeuGlnMetLeuGluSerValI 1063
QY 3726 ACCGCAAGGCTACAGATTAATAACGAAAAA----- 3759
Db 1063 eAlaGluLysGluGlnLeuLysThrAspLeuLysGluAsnIleGluMetThrIleGluAs 1083
QY 3760 -----GCGCA 3764
Db 1083 nGlnGluLeuLeuArgLeuLeuGluLysAspGluLeuLysGlnGlnGluIleValAlaGlu 1103
QY 3765 AGAAAGGATCGGAATGTAATAACAGATGATATCTTAAGAAACGATGGAAGAA 3824
Db 1103 nGluLysAsnHisAlaIleLysLysGluGluLysLeuSerArgThrCysAspArgLeuAl 1123
QY 3825 ACAAGTACAGAGATTTGATCCAAAAAAGAAATAGTAATGATCCGATGGCAATGCGG 3884
Db 1123 aGluValGlnGlu---LysLeuLysGluLysSerGlnGlnLeuGlnGluLysGlnGln 1142
QY 3885 AAATATAATTTAGTGAAGACCTGCTGTATGCCCCCTAGAAAGCAAAAGTTATG 3944
Db 1142 nLeuLeuAsnValGlnGlu-----MetSerGluMetGlnLysLysIle-- 1157
QY 3945 CGTACATTTCTGGCAATGATATGAAATTAATAAATTAACATCAACATTAATTA 4004
Db 1158 -----AsnGluIleGluAsnLeuLysAsnGluLeuLysAsnLys 1170
QY 4005 AGAAGCTTATCAATCTCGACAGACGAGAAACATTTCTCATGATTAATTAAG 4064
Db 1170 sGluLeuThrLeuGluLysMetGlnThrGluArgLeu-----GluLeuAlaGlu 1186
QY 4065 TAAGGATGTGAAGAAATACATCGATTAAGAAATTA---AAGAGGCAAAATTCCTCC 4121
Db 1186 nLysLeuAsnGluAsnTyrGlnGluValLysSerIleThrLysGluArgLysVal----- 1204
QY 4122 CGCATTTTGAATTCATGTTTACACATTTGAGAGATTAATGATTTTATTTGGA-- 4179
Db 1205 -----LeuLysGluLeuGlnLysSerPheGluThrGluArgAspHisLeuArgGlyTyr 1222

[illegible]

OY	5118	GTACAATTTGGGTCATACACTACCAAGAAAAAGAAAAAGAAATTAACAAACCTGGCA	5177
Db	1521	UGAAsnlysllelncululetyrsluysgluglnleuAenlIeIys-----	1538
OY	5178	TGGCGCAAAATATTTCTTATGAGGTCGGCGCTTGAGTGCCTATGAATATAGTTTATGA	5237
Db	1539	-----GlnlleSerGI 1542	
OY	5238	TTTAAGAGATATATCTCTAGTATGTGAATAATTGGAAAGATGAAAAACAAAGACCGAGA	5297
Db	1542	UValGlnGluSnAlaSnGlnleuLysglInPheLysGlnHlsatgLVsAlaLysAsPse	1562
OY	5298	AAATTTGAAGAAATATTTATTAACAAAAATGGACATCAGTTGGCAAGAGAAAGATAGTAC	5357
Db	1562	ralleuGlnserlleleluserlrysmetleuGlnleuThrAnAlgLeuGlnGlnserGI	1582
OY	5338	TACAGAAATCCGGGTAGTACTCCGGCAAAATTTTCTGGAACGAAATTAAGAAATGTGT	5417
Db	1582	n-GlnGlnlleGlnlleMetlleLysGlnLys-----	1592
OY	5418	GTGGACGCATATATATGCGGGGTACAAAGCTGCTAGGATGATGGAATATAGCAAAATG	5477
Db	1593	-----GlnleuMetLysaIgv 1598	
OY	5478	TGCAGAAGTGATGACATCTAAAAAATGTGGTTCTGACCTTGACATGATGATATTC	5537
Db	1598	alGlnGlnAlaLeuGlnlleGlnlAlaGAspdlInleuLysGlnsnThrLysGlnleVala	1618
OY	5538	TATGGGAAAAATCCGATTAAGTACTCGGTATCATGTTCTTCATGTTTCCGGAATG	5597
Db	1618	lalysMetLys--GlnserdlncululysglInPheLysMetThrAlaVal--	1636
OY	5598	GGGTGAAGATTTTTCGAACATTAAGAAAAAGAAATTTGGAGAAATTTGTAGGGCGCTGTA	5657
Db	1636	-----	1636
OY	5658	TGATTAATACCTTGCTGTCATATATGACATTAAGAAAGAAATGTACAGATGCTGTACACA	5717
Db	1637	-----AsnGlnThrGlnGlnLysMetCysGln-----	1645
OY	5718	ATTATAAAAAATTTATTTAGTGTAGTGAAGAACACGATGTGAAAAACA--	5763
Db	1646	-----lleGlnHlsleuLysglInPheGlnPheGlnLysleuAsnleuGln	1661
OY	5764	-----ATCAAAAAATATGGTGAATAAAGACAAAT	5795
Db	1661	uasnlleGlnThrGlnLysAsnleArGleuThrGlnlleuHlsGlnAsnleuGlnleu	1681
OY	5796	ATATTCGAGATCCTTGTCGCAAAAGATGAGAGACGCTCCGGAATATTAGACAAACA	5855
Db	1681	larGser-----ValThrLysglInArGAspAspLeuArGser--ValGlnGlnLuth	1697
OY	5856	ATTATAAAAAATTTGTGAAAAATTAAGATGGAGATTTGTGAATATTAAGTGTATGAAGATG	5915
Db	1697	rleuLys-----ValGlnArGAspGlnleuLysGlnAs	1708
OY	5916	GTCCACACAGCGATTAACCTAGTGTATAGTCAAAATATGCCCGCATCTTAGACAGATGA	5975
Db	1708	nleuArGgluthrleThr-----	1714
OY	5976	ACCAAAAGAGTGTAAAGAAAGTGTAAATTGTCAAGTGCACAGAGGCCACCGCTGTACG	6035
Db	1715	-----ArgAspLeuGlnLysglInGlnleuLyslleValAlHlsMetHlsleuLysGlnHl	1733
OY	6036	AAGGAAACACCGCTACACACAGGCTATCAGTATATCAAAAGCGACGCGATCGAAAAAGA	6095
Db	1733	sglnGlnLuthrleAspLysleuArGglYlleValSerGln-----	1746
OY	6096	AGCGAAACAGCGCGCTCAAAACAGCGCAAAAAAGGAAAAATCTTAACAACGAAT	6155
Db	1747	-----LysThrAsnGlnLileSerAsnMetGlnLysAspLeuGlnHlsSerAsnAspAlaLe	1765
OY	6156	GCGAGCA-----CAAAACGAAACCGACGACGACGACCAACAACACGAA	6200

```

Db 1765 uLysAlaGlnAspLeuLysIleGlnGluLeuArgIleAlaHisMetHisLeuLysGln 1785
OY 6201 ACAGACATCAACAGCAACAACAACAGAAATCGACGTGGCAGCAATGGTAAGGCCATTC 6260
Db 1785 uGlnGlnGluThrIle-----AspLysLeuArgGlyIleVal 1797
OY 6261 TTGGATTAACACGATAGCAGGGGGGTGATAGAGGCTTGAATCCAAAACGATAGACA 6320
Db 1797 ISeGluLysThrAspLysLeuSerAsnMetGln----- 1808
OY 6321 ATATCCTAATGGGGTGTGATGTAGTAAGCTAAAGAAATGAAATGCAATGTAT 6380
Db 1809 -----LysAspLeuGlnAsnSerAsnAla---LysIle 1818
OY 6381 GCCTCTAGAGAAAAAATTATGTATTAATATATACAAATATTAATTAATGAAATGCA 6440
Db 1818 uGlnGluLysIleGlnGluLeuLysAlaAsnGlnHisGlnLeuIleThrLeuLysLysAs 1838
OY 6441 A---AATAGCGCGACAAATGATTAAGAGCGCTTTATTAATGTCACGAATAGAAAC 6497
Db 1838 rValAsnGluThrGlnLysLysValSerGluMetGluGlnLeuLysGlnIleLysAs 1858
OY 6498 TCAATTTTGTGGTGA---AAATATATTAATGAAATCCTGCAGCAAAATGCAATGCA 6554
Db 1858 pGlnSerLeuThrLeuSerLysLeuGluIleGlnAsnLeuAsnLeuAlaGlnGluLeuHis 1878
OY 6555 AATGGAACAATTCAGATGAATTAAGAAATATGTTATACATATGTCATTAATA 6614
Db 1878 sGlnAsnLeu-----GlnGluMetLysSerValMet----- 1888
OY 6615 AGATATGTTTTGGAACTGATATTTCTAATGATTAAGAAATTAATTAAGTAAACAATAG 6674
Db 1888 ----- 1888
OY 6675 TGTAAACACCATTCATGATGAATAATAGAAAAACAGATAAAAAAAAGATAGAGA 6734
Db 1889 -----LysGlnArgAspAsnLeuArgValGlnGlu 1899
OY 6735 ATTACGTAATATTTTGGGAGAAAAATTAATAATTTATTTGGGAAGCAATATATATG 6794
Db 1899 uThrLeuLysLeuGlnArgAspGlnLeuLysGlnSerLeuGlnGluThrLysAlaArgAs 1919
OY 6795 ATTA-----ACTATCATCTCACAGACAGCAAAAGCAAAAGAA 6831
Db 1919 pLeuGlnIleGlnGlnGlnLeuLysThrAlaArgMetLeuSerLysGlnHisLysGlnThr 1939
OY 6832 -----AAATTAAGAT-----AATTAACAGTACATGACATGACCAA 6869
Db 1939 rValAspLysLeuArgGlnLysIleSerGlnLysThrIleGlnIleSerAspIleGlnIly 1959
OY 6870 A-----CTGACGCTTCCTCGAAGAGTTTGTAAAAAGGCC 6905
Db 1959 sAspLeuAspLysSerLysAspGlnLeuGlnLysLysIleGlnGlnLeuGlnLysLysGln 1979
OY 6906 C---CAATTTTGAAGATGTTCCAGAAATGGCGACAGAAATTTTGTAAAGAGAGACA 6962
Db 1979 uLeuGlnLeuLeuArg-----ValLysGlnAspValAsnMetSerHisLys 1994
OY 6963 ACACTGTGTAATTTGGAGCGCGGCTGTAAGCAATATAG-----TG 7004
Db 1994 sLysIleAsnGlnMetGlnGlnLeuLysLysGlnIlePheGlnProAsnTrpLeuCysLysC 2014
OY 7005 TTAATGCTAGTATGACGCTAGACACAGAGATGTGCAGAGCGCTGTAAACATATCAAA 7064
Db 2014 sGlnMetAspAsnIleGlnLeuThrLysLysLeuHisGlnSerLeuGlnGluIleArg-- 2033
OY 7065 TTTTATTAAAGATGGAAAAGTGAATATGAAAGCAAAAGCAAAAGTCAAAAGGATAA 7124
Db 2034 -IleValAlaLysGlnArgGlnArgGlnLeuArgGlnIleLysGlnSerLeuLysMetGlnArg 2053
OY 7125 AGATGGCAAAAAGTAAAGATATCTCTACTGAAAGACATA----- 7170

```

```

Db 2053 gAsp-----GlnPheIleAlaThrLeuArgGlnMetIleAlaArgAspArg 2068
OY 7171 -----GAGAGCGCAACATGCTCT-----CATGAATATTAA 7202
Db 2068 gGlnAsnHisGlnValLysProGlnLysArgLeuLeuSerAspGlnGlnGlnHisLeuMet 2088
OY 7203 CATGAATTTAAAGAAATTTGTGGCAATAGAGATTTGTCTGTATGCAAAAACCTTTCT 7262
Db 2088 tGlnSerLeuArgGlnLysCys-----Se 2096
OY 7263 ACAACATCAACAAAAACAACAACAATCAACATCAATCCAGTCAATATGATAGCAGAAATC 7322
Db 2096 rArgIleLysGlnLeuLeuLys-----ArgTrpSerGlnMetAspAsnHisLysGlnC 2114
OY 7323 GCTGATTTGTCTCCGAAGATTTTAACAAG---TGTAAGTGTCTGAACTTTCAAAAA 7379
Db 2114 sLeuAsnArgLeuSerLeuAspLeuGlnLysGlnIleGlnPheHisArgIleMetLysLys 2134
OY 7380 GGGATCTATGAT-----CATACAAAAAATTAATCTGAACTAAATACCTATGATTC 7433
Db 2134 sLeuLysTrpValLeuSerTrpValThrLysIleLysGlnGlnGln-----HisGlnC 2152
OY 7434 TGTAGAGAAAGCAGCATATTTATTTCTTAAGAACACAGAAAAATTAATGATATTACCT 7493
Db 2152 sIleAsnLysPheGlnMetAspPheIleAspGlnValGlnLysGlnLysGlnLeuIle 2172
OY 7494 GAAGAAAAATTTATACCTATGAG---TGTACAAAGAAAGCAAAAGTAAATTAATGATTC 7550
Db 2172 eLysIleGlnHisLysLeuGlnGlnGlnAspCysAspValProSerArgGlnLeuArgAspLeu 2192
OY 7551 GACTAATAATAATCCTTGC-----GATCCTAAGAAACCTTTGCAACCTGATTA 7598
Db 2192 sLeuAsnGlnAsnMetAspLeuHisIleGlnGlnIleLysAspPheSerGlnSerGln 2212
OY 7599 ATATATAGAAAGAAACCCCTGTGAATATAGAAATAAGAAAAATGTTTAAGATAGATTA 7658
Db 2212 uPheProSerIleLysThrGlnPheGlnGlnValLeuSerAsnArgLysGlnMetThrGln 2232
OY 7659 TGAATGAATGTTTCAAAAAATTCAAAGTCTATACAGGAGAAAAAGATAGATGTGATAC 7718
Db 2232 pPheLeuGlnGlnTrpLeuAsnThrArgPheAspIleGlnLysLeuLysAsnGlnIleGln 2252
OY 7719 TCCAAAGAACAAACATATGCTTAAGAAAT----- 7749
Db 2252 nLysGlnAsnAspArgIleCysGlnValAsnAsnPheAsnAsnArgIleIleAlaIle 2272
OY 7750 -----TTAGATGAATTAATAATGAAGACTTAAGATAGATTAATA 7790
Db 2272 eMetAsnGlnSerThrGlnPheGlnGlnArgSerAlaThrIleSerLysGlnTrpGlnGln 2292
OY 7791 TCTCCTAAAAATGCTGTGCGAAGTGCACAGAAATGAGGAATAGACATTAATAAAAACTT 7850
Db 2292 nAspLeuLysSerLeuLysGln-----LysAsnGln-----LysLeuPheLysAsnTr 2308
OY 7851 CAACCTCA-----GAGAACGGGTGGCGCAATGAATCAATATGTGTACTAT 7895
Db 2308 rGlnThrLeuLysThrSerLeuAlaSerGlnValGlnValAsnTrpThrThrGlnAspAs 2328
OY 7896 GAAATATAGTTTCCCTGATCTGGGTGACATAGTAGAGAGAAAGATGTGTTAGCAATTCG 7955
Db 2328 nLysAsnProHisValThrSerArgAlaThrGlnLeuThrThrGlnLysIleArg----- 2346
OY 7956 TGGTACTTACCTCCCTGACAAATTAATTAATATAGCTTTTGAATATATATAGGAAA 8015
Db 2347 -----GlnLeuGlnAsnSerLeuHisGlnAla----- 2355
OY 8016 ATGAGAAATTAATAAATTAAGGTAAATTAATCAACAGATGTAACAAAGTTGCTGTCTGC 8075
Db 2356 -----LysGlnSerAlaMetHisLysGlnSerLysIleIleLysMetGlnLysGln 2372
OY 8076 TTGGTGGGATGCTAATTAAGAAAGATATTTGAAAGCAATACCTGCAACACAGACAGA 8135
Db 2372 uLeuGlnValThrAsn-----AspIleIleAlaLysLeuGlnAlaLysValHisGlnLe 2390

```

QY 8136 TCGAAACCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2390 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8196 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2406 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8256 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2426 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8260 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2446 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8307 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2466 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8367 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2485 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8418 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2505 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8475 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2525 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8526 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2545 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8580 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2559 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8637 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2576 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8655 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2596 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8712 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2616 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 QY 8748 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195
 DB 2636 TAAATTCCTTTTTCAGAAAGAAATGATGATTTGAACGATTAATTAACAGAGA 8195

RESULT 15
 ID ATRX_HUMAN STANDARD; PRT: 2492 AA.
 AC P46100; P51068; Q15886; Q9NRS3; Q9H0Z1;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
 DE nuclear protein) (XNP) (Znf-HX).
 GN ATRX OR RAD54L OR XH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND

RP VARIANTS ATR-X.
 RX MEDLINE-97123494; PubMed-8968741;
 RA Picketts D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
 RA Gibbons R.J.;
 RT "ATRX encodes a novel member of the SMF2 family of proteins: mutations
 RT point to a common mechanism underlying the ATR-X syndrome.";
 RL Hum. Mol. Genet. 5:1899-1907(1996).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
 RX MEDLINE-97386582; PubMed-9244431;
 RA Villard L., Lossi A.-M., Cardoso C., Proud V., Chiaroni P.,
 RA Colliaux L., Scharitz C., Fontes M.;
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding
 RT a potential zinc finger helicase.";
 RL Genomics 43:149-155(1997).
 RN [3]
 RP SEQUENCE OF 860-2492 FROM N.A.
 RX MEDLINE-95179111; PubMed-7874112;
 RA Stayton C.L., Dabovic B., Gullisano M., Geez J., Broccoli V.,
 RA Giovanazzi S., Bossolasco M., Monaco L., Rastan S., Boncinelli E.,
 RA Bianchi M.E., Consalez G.G.;
 RT "Cloning and characterization of a new human Xq13 gene, encoding a
 RT putative helicase.";
 RL Hum. Mol. Genet. 3:1957-1964(1994).
 RN [4]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE-94214473; PubMed-8162050;
 RA Geez J., Pollard H., Consalez G., Villard L., Stayton C.L.,
 RA Millaesau P., Khrestchatsky M., Fontes M.;
 RT "Cloning and expression of the murine homologue of a putative human
 RT X-linked nuclear protein gene closely linked to PKGI in Xq13.3.";
 RL Hum. Mol. Genet. 3:39-44(1994).
 RN [5]
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RX MEDLINE-95211835; PubMed-7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RL Cell 80:837-845(1995).
 RN [6]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RA Pearce A., Chapman J.;
 RL Submitted (Dec-2000) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP E2H2 BINDING.
 RX MEDLINE-98167853; PubMed-9499421;
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
 RA Colliaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RT domain of the human E2H2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 RN [8]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE-20040663; PubMed-10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombor A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 RN [9]
 RP DISEASE.
 RX MEDLINE-20213147; PubMed-10751095;
 RA Villard L., Fontes M., Ages L.C., Geez J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 RN [10]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE-97196774; PubMed-9043863;
 RA Villard L., Lacombe D., Fontes M.;

RX VARIANT: ATR-X-S-1/V; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
CC MEDLINE=204551413; Pubmed=10995512;
CC Wada T., Kubota T., Fukushima Y., Satoh S.;
RT Molecular genetic study of Japanese patients with X-linked
RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
RL Am. J. Med. Genet. 94:242-248(2000).
CC -I- FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR, MODIFIES
CC GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
CC BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
CC -I- SUBUNIT: PROBABLY BINDS EZH2. BINDS ANNEKIN V IN A CALCIUM AND
CC PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (BY
CC similarity).
CC -I- SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
CC HETEROCHROMATIN DURING INTERPHASE AND MITOSIS. PROBABLY BY
CC INTERACTING WITH HP1.
CC -I- ALTERNATIVE PRODUCTS: 5 ISOFORMS; 1, 2, 3, 4 (SHOWN HERE) AND 5;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -I- TISSUE SPECIFICITY: UBIQUITOUS.

QY	2062	GAACCTTATGTGAAAATTTAAAAAAGA-----AAAAAAATAGCTTTC	2103
		::::: ::	
Db	55	AspMetMetGluLunSerLysGluGluGlyThrSerSerSerGluLunSerLysSerSer	74
QY	2104	AATTTGGAAAATTAATAGG-----GACTATTTAGAGAAATCAATA	2142
		::::: ::	
Db	75	GlySerSerArgSerLysArgLysProSerIleValThrLysLysValGluSerAsp	94
QY	2143	GAACCTTGTGTAGATCACTTAAAAAGAACTGCCAGATATGTMAAGACAAATTAATCAAC	2202
Db	95	GluLysProLeuAsp-----AspGluThrValAsn	104
QY	2203	GAA-----GCATGTCGAAACATCCCATTAATGCACAAACACCGTGTMAACCTCGT	2256
		::	
Db	105	GluAspAlaSerArgLunSerGluLunAsnIleThrMetGluSerLeu---ProLys	123
QY	2257	GGAGGC-----ACGGAACCCCACTAAAAATATAAAAAGAAATAGCAATACTTTAAAGG	2310
Db	124	GlyThrValIleValGluProGluProValIleuAsnGluLunPlysAspAspPheLysGly	143
QY	2311	AGTGCATACAGAGAAACAGAAATCGTGGTCTTCATAATTGAAGAAAGACACACGAA	2370
		::::	

Db 144 ProGluPheArgSerIysMetLysThrGluAsnLeuLysLysArgGlyGluAsp 163
QY 2371 GGATATATTAACGGGGGTAGAGAAAGACTTCAGACAATTATGTAGAAATATG 2430
Db 164 GlyLeuHis 166
QY 2431 ATTAACATTCATATGTAATCTTGCTTTTCAATAGGACCANTGTAGTGCAGAACGACA 2490
Db 166 166
QY 2491 GGTGATGTTATACAAACAGATTGTCGTAGAACTGATGGAGAGTGGATCCGGAAC 2550
Db 167 -----GlyLeuValSerCysThrAlaCysGlyGlnGln-----ValAsnHis 180
QY 2551 ATGCGTAAGAT-----CACGAAGAGTTATATGCTCTCTAGAACAGACAT 2598
Db 181 PheGluLysAspSerIleTyrAlaHis-----ProSerLeuGlnValLeu 195
QY 2599 ATATGTACATCCCAATTTGGACATTTACAACGGATGATCACCACCTTAAT 2649
Db 196 IleCysLysAsnLysPheLysTyrTyrMetSerAspAspIleSerArgAspSerAspGly 215
QY 2650 -----GTAATATTTGTT-----GATGATTTAGTT 2673
Db 216 MetAspGluGlnCysArgTyrPyrCysAlaGluGlyGlyAlaAsnLeuIleCysCysAspPheCys 235
QY 2674 AATAATTCCTTTTGGGGGATGTTCTTCATACGAAATATGTAAGCAACAGATATA 2733
Db 236 HisAsnAlaPheCysLysLysCysIleLeu-----ArgAsnLeuGlyArgLysGluLeu 253
QY 2734 CGAATGTATTAAGAAAGATATAC-----CTAAAGGCCCAAGAAAGTA 2778
Db 254 SerThrIleMetAspIleAsnAsnGlnTyrPyrCysTyrIleCysHisProGluProLeu 273
QY 2779 ACTGACCCAAACACAGACAGACTATCTGCGAGCTATACGTTACAGTTTGCAGATATA 2838
Db 274 LeuAsp-----LeuValThrAlaCysAsnSerVal-----PheGluAsnLeu 287
QY 2839 GGTGATATATTCGAGAGACAGATCTCTGGGAAAGAAACGGTACATGTAAAGCTGCA 2898
Db 288 -----GluGlnLeuLeuGlnGlnAsnLysLysLysIleLysValAsp 301
QY 2899 GGCATTTTGGAA-----ACGTCTTTTGTATATATACATATGCA 2937
Db 302 SerGluLysSerAsnLysValTyrGluHisThrSerArgPheSerProLysLysThrSer 321
QY 2938 CTCGAAAGCAAGGAATGTAATATATATGATGATGCC----- 2976
Db 322 SerAsnCysAsnGlyGluGlnLysLysLeuAspAspSerCysSerGlySerValThrTyr 341
QY 2977 -----CCCAATATTTA-----AAATTGAGGGA 3000
Db 342 SerTyrSerAlaLeuIleValProLysGluMetIleLysLysAlaLysLysIleGlu 361
QY 3001 AATTGGGGAACCTATAGACCAAGATGTGGGAAGCCATGAATGTATATAATAT 3060
Db 362 ThrThrAlaAsnMetLysSerSerTyrVal-----LysPhe 373
QY 3061 TTGAGAGATTAATCGGAGACCAATCAACAAAGATGATTATTCGGATATAGTATCAT 3120
Db 374 LeuLysGluAlaThrAspAsnSerGluIleSerSerAla----- 386
QY 3121 ACACCATTTGATGATTTATATCCCAAAAATTAGATGATGACCGAATGGCGCAAGATGG 3180
Db 387 -----ThrLysLeuArgGlnLeuLysAlaPheLysSerVal 398
QY 3181 TACTCGAAGTGCAGAAAAGAGATATGATTAAGTGAAGAGAGTGAAGGTGTAG 3240
Db 399 LeuAlaAspIleLysLysAlaHisLeuAlaLeuGluGlnLysLeuAsnSerGluPheArg 418
QY 3241 GATTAAGATTAATGCTACGAAAGAGAGATGCTACAGATTGTACAGATGTCACA 3300
Db 419 AlaMetAsp-----AlaValAsnLysGluLysAspThr----- 429

QY 3301 GAAGCTTTGTAATGATTAATATGATATATAGATTTATGAAGAAACAAATGATATATA 3360
Db 430 -----LysGluHisLysValIle 435
QY 3361 TCAGATTAATACAAAGA-----TTACATGAACAACGA 3393
Db 436 AspAlaLysPheGluThrLysAlaArgLysGlyLysPheCysAlaLeuGluLysLys 455
QY 3394 CAATGCTGTGTATGATATAGTATGATGATGAAGCTCCAGTACTGCGCAAAATCATATGAC 3453
Db 456 AspIleSerLysSerGluAlaLysLysLeuSerArgLysGlnValAspSerGluHisMetHis 475
QY 3454 AGAAGCTTAATGAATTTTTCGGAATTAATACCAACAAATGTGGCAAAAGTAAATAA 3513
Db 476 GlnAsnValProThrGluGlnGlnArgThrAsnLysSerThrGlyGluHisLysLys 495
QY 3514 AGTGCTACTGATGATGAATGCTGCTCATCTGATCAACACACG----- 3558
Db 496 SerAspArgLysGluGluProGlnTyrGluProAlaAsnThrSerGluAspLeuAspMet 515
QY 3559 -----TATGAATTTGTGGACATAT 3579
Db 516 AspIleValSerValProSerSerValProGluAspIlePheGluAsnLeuGluThrAla 535
QY 3580 CTCGATGATACAGAAATTTTGTATGATGATGTCAGTACAAATGAGCTTTGTATGAAAA 3639
Db 536 Met-----GluValGlnSerSerValAspHisGlnLysArgGly 548
QY 3640 AGTATGCTAAGATACGATACGAAATATATGCTTTAGACATAA-----CCACAG 3667
Db 549 SerSerLysThrGluGlnGlnGlnGlnLysSerSerValLysLeuAsnIleSerSerLys 568
QY 3688 GACCATGATGCTGCTGCTGCTGTTGTAAGAGATGATGCAACCCAGCAAG-----GTACAGATA 3744
Db 569 AspaAsnArgGlyGlyIleLysSerLysThrThrAlaLysValThrLysGluLeuTyrVal 588
QY 3745 AAA-----ACGAAAAAAAGCGAGAAAGAGATACGAGTAAACAGTG 3792
Db 589 LysLeuThrProValSerLeuProAsnSerProIleLysGlyAlaAspCysGln----- 606
QY 3793 AATGATATCTACTTAAGAAAACGATGGAAGAAACAAATAGACATTCATCCAAAAAG 3852
Db 607 -----GluValProGlnAspLysAspGlyTyrLysSerCys---GlyLeuAsnProLysLeu 624
QY 3853 AATAGTATGATATCCGATTTGCAATGCGCAATGCAATTAATTTAGTGAAGAC----- 3906
Db 625 GluLysCysGlyLeuGlnGlnLysAsnSerAspAsnGlnHisLeuValGluAsnGluVal 644
QY 3907 -----CCTGCTGTGTATGCCCCCTAGA 3930
Db 645 SerLeuLeuGluGlnLysSerAspLeuArgArgSerProArgValLysThrProLeu 664
QY 3931 AGA----- 3993
Db 665 ArgArgProThrGlnThrAsnProValThrSerAsnSerAspGluLysAsnGluThr 684
QY 3934 -----CAAAATTATGCTACATTTCTTGCAATGAT----- 3966
Db 685 ValLysGluLysGlnLysLeuSerValProValArgLysLysAspLysArgAsnSerSer 704
QY 3967 -----AATGAATTAATAAATTAATCAATCAAGTTAAT 3999
Db 705 AspSerAlaIleAspAsnProLysProAsnLysLeuProLysSerLysGlnSerGluThr 724
QY 4000 TTAAGAA-----GCTTCATCAAAATGCGACGACGAGA 4035
Db 725 ValAspGlnAsnSerAspSerAspGluMetLeuAlaIleLeuLysGlyValSerArgMet 744
QY 4036 ACATTTCTTCATGCTATTTATTAATAAGTAAAGATGCTGAAGAAATGAAGTCT----- 4089
Db 745 SerHisSerSer-----SerSerAspHisAspIleAsnGluIleHisThr 759

QY 4090 -----GATTAACAATTAAAGAGCAAAATTCCTCCCGCATTTTGG 4131
DB 760 AsnHisLysThrLeuTyrAspLeuLysThrGlnAlaGlyLysAspAspLysGlyLysArg 779
QY 4132 AGATCCAGTGTCTACACATTTGGAGATATATAGATTTTATTGGACAGATATATCA 4191
DB 760 LysArgLysSerSerThrSerGlySer-----AspPhe-----AspThrLys 793
QY 4192 AAAGTCACTGGTGAGGA-----AGTAACTAAAGAGCAAAATGATTTCTT 4239
DB 794 LysGlyLysSerAlaLysSerSerLelSerLysLysArgGlnThrGlnSerGlu 813
QY 4240 TTCAAAATGGTGAACCAAAATCTCTTAATGAAAAACACCCCAAGATGGTGACAGAA 4299
DB 814 SerSerAsnTyrAspSerGlu-----820
QY 4300 CATAGTCATGAGATATGAGACTATGCTATGCTACTAGTAAATTTGGGCAAAAAA 4359
DB 821 -----LeuGlnLysGlnLelLysSerMetSerLysLelGlyAlaAlaArg 835
QY 4360 -----GATGATTT-----4368
DB 836 ThrThrLysLysAlaGlyLelProAsnThrLysAspPheAspSerSerGluAspGlnLysHis 855
QY 4369 -----ACCGAAACTACGCTTACACACAGCTCAAAATTTAGT-----4404
DB 856 SerLysLysGlyMetAspAsnGlnGlyHisLysAsnLeuLysThrSerGlnGlnLysSer 875
QY 4404 -----4404
DB 876 SerAspAspAlaGlnArgLysGlnGlnArgGlnThrPheSerSerAlaGlnGlyThrVal 895
QY 4405 GACAAAGACACCACTTTGGAGAAATTTGCCAANCA-----CCCAAGTTTATACGA 4455
DB 896 AspLysAspPheThrLelMetGlnLelArgAspArgLeuProLysLysGlnGlnAlaSer 915
QY 4456 TGCTTACCGAATGTGACGACACTATTCCTATACACAGCAAAATAT-----4503
DB 916 AlaSerThrAspLysValAspLysLeuSerGlyLysGlnGlnSerPheThrSerLeuGln 935
QY 4504 TTGAGAGATGTGCAGAAAAATTAAGTCAATGACCAATTTAGTGTATACAGAAATG 4563
DB 936 ValArgLysValAlaGlnThrLysGlnLysSerLysHisLeuLys-----ThrLysThr 953
QY 4564 AATTAAGAAATGGAGGACTACGTT-----AAATATATGAAAAAATAA 4608
DB 954 CysLysLysValAlaGlnAspLysLeuSerAspLelAlaGlnLysPheLeuLysLysAspGln 973
QY 4609 GAGTGTATCCACAGATTAATATTTACAAAGATGAACGCGACAAAAAAGATTGATAGA 4668
DB 974 -----SerAspLysThrSerGluAsp-----AspLysLysGln-----984
QY 4669 CAACACATTGTGTATGTTAGTACATACTAGAACGAATGCAACAGATTACTTGAAAC 4728
DB 984 -----984
QY 4729 AGGAAATTTACTGCTAGTTGGTGATTAAGCTTGGAAGTGCCTCTGTGTACAAAGAAAT 4788
DB 985 -----SerLysLysGlyThrGln-----990
QY 4789 ATACAATTTAGAAAAACAGGCTTACTATGATGCCGACAAACATTTGGGTGCCAAAAA 4848
DB 991 -----GlnLysLysLysProSerAspPheLysLys-----Lys 1001
QY 4849 TTTATGAAAAATGACGACAAATATATCTAACATTTTCGAGTAAGATGAAGTGAAGATTA 4908
DB 1002 ValLelLysMetGlnGlnGlnLysThrGlnSerSer-----1013
QY 4909 GTAAGAGGGAACACAGCTGCTATTATAGTGGCAAAACAAAGTCTTAATATACAT 4968
DB 1014 -----AspGlyThrGln-----1017
QY 4969 AACTTGAAGAAATGACATGAGATGTGCTTTTCTCTCTGCTGACATGATATATGTTT 5028

DB 1018 LysLeuProGlnArgGlnGlnLelCysHisPheProLysGlyLelLysGlnLelLysAsn 1037
QY 5029 CAGCATTTGGATGGCAATTTATACAGATCCAGAGTTAAAGTGAATAATGGTTGCCAAA 5088
DB 1038 GlyThrThrAspGlnGlnLysLysSerLysLysLelArgAspLysThrSerLysLys 1057
QY 5089 AGATTGATG-----GAATGGCGGACACGGAAGGTTACAAATTTG-----5127
DB 1058 AspGlnLeuSerAspTyrAlaGlnLysSerThrGlyLysGlyLysPheCysAspSerSer 1077
QY 5128 -----GGTCACTACTACAAAAAATAA-----5151
DB 1078 GlnAspLysLysSerLysAsnGlnAlaTyrGlyArgGlnLysLysArgCysLysLeuLeu 1097
QY 5152 -----GAAAAAGAAAAATAAAGCAAGTGGAGCGCCACAAATTTCTAT 5196
DB 1098 GlyLysSerSerArgLysArgGlnAspCysSerSerSerAspThrClnLysLysMet 1117
QY 5197 GAGGTCCCGCTGTAGTCTATGAATATATGATTTTATGATTTAAGATATAATCTA 5256
DB 1118 LysGlnAspLysCysAsnSerSer-----AspLysArg-----LeuLys 1130
QY 5257 GGTATTTGATTAATTTGGAAGATGAAAAACAAAGACCGAGAAATTTGAAGAAATATTT 5316
DB 1131 ArgLelGlnLelArgGlnArgLysAsnLeuSerSerLysArgAsnThrLysGlnLel 1149
QY 5317 AACAAAAATGACACATGCTAGTTGGCAAGAGAGTATGATCTACTACAGAAAT-----5367
DB 1150 -----GlnSerGlySerSerSerSerAspAlaGlnGlnSerSerGlnAspAsnLysLysLys 1168
QY 5368 -----CCCGAGTACTGCGCAAAATTTTCTGCAACGAAAAATTAAGAAATGTGTG 5418
DB 1169 LysGlnArgThrSerSerLysLysLysAlaValLelValLysGlnLysLysArg-----1186
QY 5419 TGGAAACGAATGATATGCGGTGACAAACGTGAGGATGATGAATAGTGAATAGT 5478
DB 1187 -----AsnSerLeuArgThrSerThrLysArgLysGlnAlaSerPheThrSerSerSer 1205
QY 5479 GCAAGA-----AGTGAATGAAGATCTAAAAAATGTGTTCTGCTACTTACATGATGAT 5532
DB 1206 SerAspLelGlnLysAspAspGlnAsnSerLelLelGlnLysSerSerAspGlnLys 1225
QY 5533 ---TATCCTATGCGCAAAATTCGCAATGAAGTACTGCTATCAGTTTCTTCGATGGTT 5589
DB 1226 LelLysProValThrGlnAsn-----LeuValLeuSer 1236
QY 5590 GCCGAATGGGTGAAGATTTTTCGAAACATTA-----GAAAGCAATTTGAAGAAATTTGTA 5646
DB 1237 SerHisThrGly-----PheCysGlnSerSerGlyAspGlnAlaLeuSerLysSerVal 1254
QY 5647 GGGCGTGTATGATTTACTTGTGATATGAGATGAAGATGAAGAAAGAAATGTACAGAT 5706
DB 1255 ProValThrValAspAspAspAspAspAspAspProGlnAsnArgLelAla-----1272
QY 5707 GCGTGTACAAATATTAATAATTTATTTAGTGAAGAAACACAGATGAAAAACAAATC 5766
DB 1273 -----LysLysMetLeuGln-----1278
QY 5767 AAAAAATATGTGAGATAAAGACAAATAATATATTCGAGATCCTGTGCCAAAGTGA 5826
DB 1279 -----GlnLelLysAlaAsnLeuSerSerAspGlnAspLysSerSerAsp-----1293
QY 5827 GAGAGCGCTCGCAATATTTAGCAAAACAAATTAATAAATTTTGTGAATAATAAGTGA 5886
DB 1294 ---AspGlnProGlnGlnLysLysArgThrGlnGlnLysGlnAsnGlnLysAsnProGly 1312
QY 5887 GATTGTGAATATATAGCTATGAAGATGTGTCCACACCGCATTAACGTAGTAAATAGT 5946
DB 1313 AspGlnGlnAlaLys-----AsnGln 1319
QY 5947 CAATAATGCCGCAATCATTAACAGATGAACCAAAAGAAAGTTGAAGAAAGTGTATATGT 6006

Db 1320 ValAsnSerGluSerAspSerGlu-----GluSerLys----- 1332
QY 6007 CAATGCCACGAGGTCCACCAGCTGTACAGAGGAAACCGCTCACCGGATACATG 6066
Db 1333 -----LysProArgTyrArgHisArgLeu-----Leu 1341
QY 6067 ATATCAAAAGCGACGATCGAAA-----AAGAGCGAAAACAGCGCGCGCT 6114
Db 1342 ArgHisLysLeuThrValSerAspGlyGluSerGlyGluGluLysThrLysPro--- 1360
QY 6115 ACAAAACACCGCAAAAATGTGCAAAATTAACACAGAAATGCGACACAAACACGAC 6174
Db 1361 ---LysGluHisLysGluValLysGly----- 1368
QY 6175 CGACGACGACGACACAAACAGCAATCAACAGCAACAAACAGCAATGTCAC 6234
Db 1369 -----ArgAsnArgArgLysValSerSerGluAspSerGluAspSerAsp 1383
QY 6235 GTG---GGCAACATGGTAAAGCCATTCTTCGAATTAACAGATAGCAGGGGTGAAAT 6291
Db 1384 PheGlnGluSerGlyValSerGluValSerGluSerGluAspGluGln----- 1400
QY 6292 GAGGGTTGTAATCCAAAAGATGACATATCTAAATGGGCTGTATTTAGGTAAAG 6351
Db 1401 -----ArgProArgThrArgSerAlaLysLys-----AlaGlu 1411
QY 6352 TCTAAAGAAATGAAAATGGCATATGTATGCTCTCAGAGAAAATTAATGTATTAAT 6411
Db 1412 LeuGlnGluAsnGlnArgSerTyrLysGlnLysLysArgArgLysLysValGln 1431
QY 6412 AATATCAATTTAATATATGAAATGAAATTAACCGTGACATATTAAGAGGCT 6471
Db 1432 Glu-----AspSerSerSerGluAsnLysSerAsnSerGluGluGluGlu 1447
QY 6472 TTTATTAATGTCCACCAATTAATGCTCAATTTTGTGTTAAATATATTAATTAAT 6531
Db 1448 GluLysGluGluGluGluGluGluGluGlu-----GluGluGlu 1460
QY 6532 CCTGCACGCAAAATGAAATGCAAAATGCAAAATGCAAAATGCAAAATGCAAAATG 6591
Db 1461 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1480
QY 6592 TATATACATATGATGATTAATAAGATATGTTTGGAACTGATTTCTATGATTA 6651
Db 1481 -----ArgLysLysLeuLysAsp 1487
QY 6652 AAAATTAATCTGTACAAATAGTGTACAAACCATTCATGAATATAATAGAAAAA 6711
Db 1488 LysLeuArgThrGluThrGlnAsnAlaLeu----- 1497
QY 6712 CAGGATAAAAAAGATGAGAAATTAATTAATTTGGGAGAAAAATTAATTAATTT 6771
Db 1498 -----LysGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 1510
QY 6772 ATTTGGGAAGATGATATATGATTAATCACTATCATCTCAGACAGAAAAAGAA 6831
Db 1511 -----GluArgGluArgGlu 1515
QY 6832 AAAATTAAGATTAATTAACGATCAATGAC-----ATGACC 6867
Db 1516 LysLeuArgGluValIleGluIleGluAspLysSerProThrLysCysProIleThr 1535
QY 6868 AAACTGACGCTTCCCTTGAAGAGTTTGTAAAAAGGCC----- 6906
Db 1536 LysLeuValLeuAspLysGluGluThrLysGluProLeuValGlnValHisArgAsn 1555
QY 6907 -----CAATTTTGGATGTTGTTCA 6927
Db 1556 MetValIleLysLysProHisGlnValAspGlyValGlnPheMet----- 1571
QY 6928 GAATGGCCAGACAAATTTGTATAAGAGCAAGACAGTTGTTAAATTTGAGCGCGC 6987
Db 1572 ---Trip---AspCysCysGluSerValLysLysThrLysLysSerProGlySerGly 1589

QY 6988 TGTAAAGATATGATGTATAGTACTATGACGGTAAACAGACAGAAATGTCAGAG--- 7044
Db 1590 CysIleLeuAlaHisCysMetGlyLeu-----GlyLysThrLeuGlnValLysPhe 1607
QY 7045 -----GCGTGTACATAT 7059
Db 1608 LeuHisThrValLeuLeuLysAspLysLeuAspPheSerThrAlaLeuValCysPro 1627
QY 7060 CAAAATTTTATTAAGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 7119
Db 1628 LeuAsnThrAlaLeuAsnTrpMetAsnGluPheGlnLysTrpGlnGluLysAsp 1647
QY 7120 GATTAAGATGGCAAAAAGTAAAGATTAATCTTCTTACTGAAAGACATAGAAAGCA 7179
Db 1648 AspGlu----- 1649
QY 7180 ACATGTGCTCATGATATTTAAACATGAATTAAGAAATATGTGCAATTAAGATTTG 7239
Db 1650 -----LysLeuGluValSerGluLeuAlaThrValLys----- 1660
QY 7240 TCTTGTATGCAAAAACCTTCTTCACAACATCAAAAACACACAAATCAATCATC 7299
Db 1660 ----- 1660
QY 7300 GATGTAAATGATATGCCAGATGCTGATTAATGTCTCGAAGAAATTAACAAGTGTAG 7359
Db 1661 -----ArgProGlnGluArgSerTyrMetLeuGlnArgTrpGlnGlu----- 1674
QY 7360 TGTCCGAACTTCAAAAAGGATCTATGATTCATACAAAAAATTAATGTAACCTTAA 7419
Db 1675 -----AspGlyLysValMetIle----- 1680
QY 7420 ATACATTAATGTGTAGAGAAAGCAGCATATTAATTAATTAAGAGAGAGAAATTAAT 7479
Db 1681 -----IleGlyTyrGluMetTyrArgAsnLeuAlaGlnGlyArgAsnVal 1695
QY 7480 ATGATATTAATCTTGAAGAAAAATTTATACCTATTTAGTCTACAAAGAAAGAAAGT 7539
Db 1696 LysSerArgLysLeuLysGluIlePhe----- 1704
QY 7540 AAAATAGTGGATATATATATTCCTGACCTTAAGAAACCTTATGACCGTAA 7599
Db 1705 -----AsnLysAlaLeuValAspPro-----GlyProAsp--- 1714
QY 7600 TATATAGAGAGAAACCTTGTGAATAATGAGAGAAATGCTTTTAAGATGATTA 7659
Db 1715 -----PheValValCysAsp 1719
QY 7660 GAATGGAATGTTACAAAATTTCAAACTTATCAGAGAAAAAAGATATGTCCT 7719
Db 1720 GluGlnHisLysLeuLysAsnGlnAlaSerAlaValSerLysAlaMetAsnSerIleArg 1739
QY 7720 CCAAGAA-----GAAACATATG 7737
Db 1740 SerArgArgArgIleLeuThrGlyThrProLeuGlnAsnAsnLeuIleGluTyrHis 1759
QY 7738 TGCCTTAAGCAATTTAGATGAATTAATTAAGAAAGCTTAAGATGATTAATTAATTA 7797
Db 1760 CysMetAlaAsnHe-----IleLysGluAsn-----LeuLeu 1770
QY 7798 AAAATGCTTCGTCGAACTGACAGAAATGAAGATTAATTAATTAATTAATTAATTA 7857
Db 1771 GlySerIleLysGlu---PheArgAsnArgPheIleAsnProIleGlnAsn----- 1786
QY 7858 GAGAGCGGTGGCAATGATCAATATGAT-----ACTATGAA----- 7899
Db 1787 ---GlyGlnCysAlaAspSerThrMetValAspValArgValLysLysArgAlaHis 1805
QY 7900 TATAGTTCCGTATCTGCTGATGATAGAGAGAAAGATATGTTACAAATTTGATGTGT 7959
Db 1806 IleLeuTyrGluMetLeuAlaGlnLysValGlnArgLysAspTyrThrAlaLeuThrLys 1825

```

QY 7960 TACTTACCTCCC-----GTAGAAATAA 7983.
    :|||:
Db 1826 PheLeuProPolysHisGluTyrValLeuAlaValArgMetThrSerIleGlnCysLys 1845
QY 7984 TTATATAGGTTTT---GAATACATATATGAAATGAGAAATATAAATTAAGGTGA 8040
    :|||:
Db 1846 LeuTyrGlnTyrTyrLeuAspHisLeuThrGlyValGlyAsnAspSerCysGluGlyArg 1865
QY 8041 AATAA-----TACAAAGATGTAACAACGTTTCGTTCTGTTG----- 8079
    :|||:
Db 1866 GlyLysAlaGlyAlaValLeuPheGlnAspPheGlnMetLeuSerArgIleThrPheHis 1885
QY 8080 ---TEGGATGCTAATAGAAAAGATATTTGAAAGCAATAGCTGCAAGCACCAGACAGAT 8136
    :|||:
Db 1886 ProIrpCysLeuGlnLeuAspTyrIleSerLys-----GluAsn 1898
QY 8137 GCAAAACCTTTTAGAAAAGAAAGATGATGATTTGAACGCATATATATACAAAT 8196
    :|||:
Db 1899 LysGlyTyrPheAspGluAspSerMetAspGluPhe-----IleAlaSerAsp 1914
QY 8197 AAGTGTGACATTAAGACGATCCACCTGTGTGATTTATATCTCAACGTTTCGATGG 8256
    :|||:
Db 1915 SerAspGluThrSerMetSerLeuSerSerAspAspTyrThrLysLys----- 1931
QY 8257 ATGACTGAATGCTGCAATATATTATTGTAAGCATGATGGAAGAAATGGAATTTAA 8316
    :|||:
Db 1932 -----LysLysGlyLysLysGlyLys 1938
QY 8317 AATCATGTGATCACTGTAACATCTGACAGATGCAAGAAATGAT----- 8361
    :|||:
Db 1939 Lys-----AspSerSerSerSerLysSerGlySerAspAspValGluValIleLys 1956
QY 8362 -----TATGATGAAAT----- 8373
Db 1957 ValTrpAsnSerArgSerArgGlyGlyGlyGluGlyAsnValAspGluThrGlyAsnAsn 1976
QY 8374 -----AAGTGCACACGTCGTAAACGACATGCAAGAA----- 8406
Db 1977 ProSerValSerLeuLysLeuGlnSerLysAlaThrSerSerSerAsnProSerSer 1996
QY 8407 -----TATAAATATTTGTTTAAATGGAATCTCTATTCGATATACA 8451
    :|||:
Db 1997 ProAlaProAspTyrTyrLysAspPheValThrAspAlaAspAlaGluValLeuGlnHis 2016
QY 8452 TCATAATTAATACAAAGAAATGTATGAA-----CAACCAATATATACAAA 8496
    :|||:
Db 2017 SerGlyLysMetValLeuLeuPheGluIleLeuArgMetAlaGluIleGlyAspLys 2036
QY 8497 ATCTCTACTTAT-----GATCATGTCGCAAAATTTTGTACAAAAG 8535
    :|||:
Db 2037 ValLeuValPheSerGlnSerLeuIleSerLeuAspLeuIleGluAspPheLeuGlnLeu 2056
QY 8536 TTGAACACTTTTAAAGTGAATGTTCTGTGAGACTTTTGAATATCTCATGAACA 8595
    :|||:
Db 2057 AlaSerArgGluLysThrGlu-----AspLys 2065
QY 8596 ACTAAGTGTGAATTAATTAATTAATGAAGATGTTCTTCCATATACGAACATAT 8655
    :|||:
Db 2066 AspLysProLeuIleTyrLys-----GlyGluGlyLysTyrLeuArgAsnIleAspTyrTyr 2084
QY 8656 GCTTCGAAGAAACACCAAAAGTTATAAGAGCTGCAGTTGTACACTACTCTCTAAG 8715
    :|||:
Db 2085 Arg----- 2085
QY 8716 AATCATGTGATTAATGCTCTACGATCAAAAC---AAGATGATGTGAAGAAATTAACA 8772
    :|||:
Db 2086 -----LeuAspGlySerThrThrAlaGlnSerArgLysTyrPalagLunPheAsn 2103
QY 8773 ACTTTTACTCTGCTCGAAGATGATTTATGATTAATATCTGATTAATGGAACGATAC 8832
    :|||:
Db 2104 AspGluThr-AsnValArgGlyArgLeuPheIleIleSerThrLysAlaGlySerLeuGln 2123
QY 8833 CTTCGTTCTTAATA----- 8845

```

```

Db 2123 yIleAsnLeuValAlaAlaAsnArgValIleIlePheAspAlaSerTrpAsnProSerTyr 2143
    :|||:
QY 8846 -----GTTGAGATGATTAACAAGGTGATTTGATTCCTCCACGAAGACATTTATGA-- 8899
    :|||:
Db 2143 RasPleGlnSerIlePheArgValIlyArgPheGlnGlnThrLysProValIlyValTyr 2163
QY 8900 -----CA 8901
Db 2163 rArgPheLeuAlaGlnGlyThrMetGluAspLysIleTyrAspArgGlnValThrLysGln 2183
QY 8902 AGACCTATCACTGCATATATTAATATAGAAAAGCTGATTAAGAAATTTTA----- 8949
Db 2183 nSerLeuSerPheArgValValAspGlnGlnGlnValGlnArgHisPheThrMetAsnGln 2203
QY 8950 -----AA 8951
Db 2203 uLeuThrGluLeuTyrThrPheGluProAspLeuLeuAspAspProAsnSerGluLysLys 2223
QY 8952 AAAAAAACTTCTTACTTCTGCTTCAGTCAAGCAAAATTTGAGTCA----- 9000
    :|||:
Db 2223 sLysLysArgAspThrProMetLeuProLysAspThrIleLeuAlaGluLeuGlnIle 2243
QY 9001 -----AATATTAATCGGA 9014
Db 2243 eHisLysGlnHisIleValGlyTyrHisGlnHisAspSerLeuLeuAspHisLysGluGln 2263
QY 9015 AGAAGAGTTGTGCTTTGAGCAATGAATATATGATTCAGATTTATTC----- 9063
    :|||:
Db 2263 uGlnGluLeuThrGlnGluGlnGluArgGlyAlaIleThrPalagLunTyrGluAlaGluLysLys 2283
QY 9064 -----GATATTAATTAAGCAACTATATGATGCACTTCATATTC 9104
    :|||:
Db 2283 sValLeuThrMetArgPheAsnIleProThrGlyThrAsnLeuProValSerPheAs 2303
QY 9105 TGAATAAAT----- 9114
    :|||:
Db 2303 nSerGlnThrProTyrIleProPheAsnLeuGlyAlaLeuSerAlaMetSerAsnGlnGln 2323
QY 9115 -----AAAAATATTTGAAACATCAATGAAAGCAAC 9146
    :|||:
Db 2323 nLeuGlnAspLeuIleAsnGlnGlyArgGluLysValValGluAlaThrAsnSerValThr 2343
QY 9147 CGAAATCGT-----AAAACATGCTGGGAAAT 9174
    :|||:
Db 2343 rAlaValArgIleGlnProLeuGlnAspIleIleSerAlaValThrLysGlnAsn 2361

```

Search completed: April 28, 2003, 11:32:25
 Job time : 803 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: April 28, 2003, 11:02:12 ; Search time 296 Seconds
(without alignments)
6903.491 Million cell updates/sec

Title: US-10-087-013-1

Perfect score: 19455
Sequence: 1 atgggggtctctctgaata.....atatatgatgaagatattg 10628

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 283224 segs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+ n2p model -DEV-xlp
-O-/cgn2.1/USPTO.spool/US10087013.rpr.rpt.28042003.102846.20277/app.query.fasta.1.10823
-DB-pir.73 -OFMT-fasta -SUFFIX-n2p -MINMATCH-0.1 -IOOPT-0 -IOPEXT-0
-UNITS-bits -START-1 -END-1 -MATRIX-blosum62 -TRANS-numa40.cdi -LIST-45
-DOCALLGN-200 -THR_SCORE-pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15 -MODE-LOCAL
-OUTFMT-pto -NOR-ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000
-USER=US10087013.qcon1.1.520.etrnat.28042003.102846.20277 -NCPU-6 -ICPU-3
-NO_XLPHY -NO_MMAP -LARGEQUERY -NEG_SCORES-0 -WAIT -LONGLOG -DEV_TIMEOUT-120
-WARN_TIMEOUT-30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6 -FGAPEXT-7
-YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database:

1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4515.5	23.2	2706	2	T28155	variant-specific s
2	4260	21.9	3006	2	T28625	variant-specific s
3	4010	20.6	3026	2	T28431	variant surface pr
4	3501	18.0	3078	2	T28432	variant-specific s
5	3154	16.2	2924	2	T18378	variant-specific s
6	2727	14.0	2647	2	T28161	hypothetical prote
7	2726.5	14.0	2647	2	T28626	variant-specific s
8	2678.5	13.8	2212	2	T28157	erythrocyte membra
9	2303	11.8	2135	2	T14602	variant-specific s
10	1773	9.1	2197	2	B71600	variant-specific s
11	1755	9.0	2182	2	T28634	variant-specific s
12	1731.5	8.9	2228	2	T14029	variant-specific s
13	1586	8.2	1711	2	C71625	variant-specific s
14	1542.5	7.9	2042	2	T18399	variant-specific s

15	1249	6.4	1729	2	T18396	erythrocyte membra
16	835.5	4.3	431	2	F71600	variant-specific s
17	618	3.2	1435	2	A37793	erythrocyte-bindin
18	561.5	2.9	4550	2	T18440	hypothetical prote
19	558	2.9	3724	2	T18427	hypothetical prote
20	539.5	2.8	440	2	E71625	variant-specific s
21	535.5	2.8	3394	2	T18501	hypothetical prote
22	513.5	2.6	4981	2	T18489	hypothetical prote
23	461	2.4	2523	2	T18402	hypothetical prote
24	453.5	2.3	3844	2	T18477	asparagine/asparta
25	450.5	2.3	3973	2	B71612	hypothetical prote
26	443.5	2.3	2401	2	T28676	rhodopy protein
27	435	2.2	2829	2	A42771	reticulocyte-bind
28	434	2.2	2510	2	T28160	hypothetical prote
29	427.5	2.2	2657	2	T18497	hypothetical prote
30	427	2.2	1979	2	C71622	hypothetical prote
31	424.5	2.2	2437	2	T18482	hypothetical prote
32	420.5	2.2	6669	2	S55024	nebulin, skeletal
33	409	2.2	2269	2	T28677	rhodopy protein
34	409	2.1	248	2	C71624	variant-specific s
35	404	2.1	2573	2	D71614	hypothetical prote
36	392	2.0	2485	1	H71621	serine/threonine-s
37	387	2.0	4688	2	F82885	hypothetical prote
38	384.5	2.0	3135	2	A48584	transmission block
39	366.5	1.9	1939	2	T18372	repeat organellar
40	366	1.9	2295	2	B71621	probable membrane
41	362	1.9	1819	2	A71928	cag island protein
42	362	1.9	2539	2	B71619	hypothetical prote
43	357.5	1.8	2500	2	G71609	hypothetical prote
44	355	1.8	1927	2	G64585	cag pathogenicity
45	354.5	1.8	1045	2	T18373	erythrocyte bindin

ALIGNMENTS

RESULT 1

T28155 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragme
N:Alternate names: erythrocyte membrane binding protein 1 (EMBP1)
C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T28155

R:Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.

Nature 388, 292-295, 1997

A:Title: Plasmodium falciparum rosetting is mediated by PfEMP1 and requires complemen

A:Reference number: Z20477; PMID:97373957; PMID:9230440

A:Accession: T28155

A:Status: preliminary; translated from GB/EMBL/DBD

A:Molecule type: DNA

A:Residues: 1-2706 <ROW>

A:Cross-references: EMBL:Y13402; PIDN:CAA73831.1

A:Experimental source: strain IT 4/25/5

C:Genetics:

A:Introns: 2493/3

A>Note: R29R+var1

Alignment Scores:	Pred. No.:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
Score:	1.19e-214	2706	1191	372	871	1315	123
Percent Similarity:	4515.50						
Best Local Similarity:	41.69%						
Query Match:	31.77%						
DB:	23.21%						
US-10-087-013-1 (1-10628) x T28155 (1-2706)							
OY 100 AGTCAAGAAATGTTTGGACGTTATGCCAAAATATAAGACATCCATCAAAA-----	153						
DB 14 SerAlaThrAspValLeuGluTyrIleAlaThrCylTyrAsnGlnGluTyrSGluTyr	33						
OY 154 ---TATCAAGAAAGAAATGATGATTCGTTGAAGGAGGATTTGACAAAGCAAGATTT---	207						
DB 34 ValTyrProTyrGlu-----AsnGluLeuTyrGlyIleLeuSerAsnAlaIlePheVal	51						

QY 208 -----CGTGGTCTCTTCTACGCCAGTAATAG 237
Db 52 AspGlnLeuArgLysGlnLeuAsnIleGluSerProGlyProSerAsp----- 67
QY 238 CATATAATATATATATCATATCCATGTAATTTAGATCATAGAACATCTACTAATTTACGG 297
Db 68 -----SerCysSerLeuAspHisLysPheHisThrAsnIleAsn 80
QY 298 TATGATGATGTGAATTTAGACATCTTGGCATGGTAGACAAAACGATTTGATGAA 357
Db 81 ThrGlnThrThrGlnGlnArgLysProCysTyrGlnArgLysGlnLysAspGlnPheSerAsn 100
QY 358 GATGAGAAATCTGAATGTGA---AATAAATACGTATATTAAGAAAGAAATGATGCT 414
Db 101 GlnGlnGlnLysLysCysGlnSerAspLysIleArgAspLysGlnIleLysSerIleGln 120
QY 415 ATAGCCCTGCGCCACCTAGAACAGCATATGTGTGATTAAGAAACCTTGAAAGCTCTAAT 474
Db 121 GlnAlaCysAlaLeuProPheArgGlnAsnLeuCysAspArgAsnLeuGlnLysIle 140
QY 475 GATATATAATCCCAAAATATTCATGATTTATTTGGAATGTACTAGTACGCAAAATAC 534
Db 141 AsnLysAsnThrAsnThrThrHisAspLeuGlnLysAsnValLeuValThrAlaLysLys 160
QY 535 GAAGGTGAATCATGTGTATATATCATCCACATAAAGAACT-----TCAGAC 582
Db 161 GlnGlnLysSerIleValAsnAsnHisProAspLysAsnSerSerGlnLysSerSer 180
QY 583 GCTTGTACTGCTCTTGCACAGATTTTGCAGATATAGTATGTATGTAGAGAGAAATAG 642
Db 181 IleCysThrAlaLeuAlaArgSerPheAlaAspIleGlnLysIleValArgGlnArgAsp 200
QY 643 AGTTTAAACCAAAATCTCCATGACAAAGTAGAAACGGCTCTCGAAGGTTTTCAGAAA 702
Db 201 MetPheLysProAsnAspAlaAspLysValGlnLysGlnLeuGlnValAlaPheGlnLys 220
QY 703 ATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 759
Db 221 IleTyrAsnSerLeuProSerProAlaGlnLysHisLysThrAlaHisAspAspLysSerGly 240
QY 760 AATTTATTAATTAATTAAGAGAACAGCTGGAATGTGAATGAATGAATGAATGAATGAAT 819
Db 241 AsnTyrTyrLysLeuAlaArgLysAspTrpTrpAlaIleAsnArgLysGlnValIleTrpLysAla 260
QY 820 ATTAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 879
Db 261 IleThrCysArgAlaProAsnGlnLysIleAsnPheArgAsnIleSerGlnLysMetLys 280
QY 880 TTATTTT---TCAAAATCTTAATATGATGCGCCATTAACAGAAAGGTTCTACCAATTTAGAT 936
Db 281 AlaPheThrSerGlnGlnLysCysGlnHisSerGlnIleHisAsnValIleProThrAsnLeuAsp 300
QY 937 TATGTCCTCATATTTACGTGGTTCGACGCAATGGGAGAAAGAGTTTGGCCGAAAAAGA 996
Db 301 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 1000
QY 997 AATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1056
Db 321 LysIleLysLysGlnLysAsnValLysLysLysLysLysLysLysLysLysLysLysLysLys 340
QY 1057 AGTCATTAATGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1116
Db 341 SerGlnLysArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 360
QY 1117 AATAAGTCTACTGACTGCTTCCATTAATGCAAAAGTTTGTGAAGTTTGGTTAGGCAATCA 1176
Db 361 LeuAsnCysProArgCysGlnAsnAlaCysSerAsnTyrTrpIleGlnIleGln 380
QY 1177 CAAGAGAGATTTTAAACCAAAAGAAAGAAATATGAAAAAGAAATATTTATTCG 1236
Db 381 ArgLysGlnPheAspLysGlnLysArgLysLysLysLysLysLysLysLysLysLysLysLys 395

QY 1237 AACGATTAACAAATTTTGCATTAATTT---AATAGCAATATTTAATTAACAAATTTATGAA 1293
Db 396 -----LysIleLysThrAsnIleSerAsnAsnGlnAsnAspLysGlnPheLysGln 412
QY 1294 AAACCTTAAGGAAACGCAATATGCAATTAATGACACTTTTAAATTTAATTAATGAAGA 1353
Db 413 AsnLeuAspLysLysGlnLysSerThrIleAsnThrPheLysGlnSerLeuAsnHisGln 432
QY 1354 AAGTATTTGAAGAGATTTACAGAGAAAGAAAGATTTACTTTTACTTAACAGTCTGAT 1413
Db 433 LysGlnCysGlnHisAsnIleAspLysLysAsnLysThrAsnPheLysAsnHisLeuGln 452
QY 1414 GACAAAGGATTTTATGCTCAGAAATTTGCAAGTGTGCGGACGTCCGAGCGGGTCAAA 1473
Db 453 -----ThrPheLysProSerGlnLysGlnLysLysLysLysLysLysLysLysLys 469
QY 1474 TGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1533
Db 470 CysSerAsnGlnLysCysThrProValThrGlnAsnGlnLysPhe-----AsnSer 485
QY 1534 GACTTAAACCTCCATGCG-----GGTGTGAAGCTTACTTAATTAATCACTGTC 1578
Db 486 AsnAsnArgLeuProThrAspThrSerThrLysAsnLeuAsnAlaThrAsnIleAspMet 505
QY 1579 CTTTATAGTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1638
Db 506 LeuValAsnAspGlnLysGlnLysAsnAlaIleAspAsnGlnLysLysLysLysLysLys 525
QY 1639 TCAACTATTTTCAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1698
Db 526 TyrGlnLysLysLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 544
QY 1699 ATTAATAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1749
Db 545 IleAspGlnCysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 562
QY 1750 AAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1809
Db 563 -----IleAlaPheAsnValLeuPheGlnArgTrpLeuArgLysPheValArgAspHis 580
QY 1810 ATTAAGTGAATGACAAACTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1860
Db 581 AsnArgLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 600
QY 1861 TGTATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
Db 601 CysIleLysArgCysLysThrAsnLysGlnLysValGlnLysLysLysLysLysLysLys 620
QY 1921 GAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1971
Db 621 AlaGlnTrpAspLysLysLysGlnHisLysLysLysLysLysLysLysLysLysLysLys 640
QY 1972 -----CAATGCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2006
Db 641 IleProTyrTrpIleThrGlnLysPheLysLysLysLysLysLysLysLysLysLysLys 658
QY 2017 TTTTAAAGTTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2076
Db 659 PheLysAlaLeuGlnLysValAspThrIleAsn----- 669
QY 2077 ATTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2136
Db 669 ----- 669
QY 2137 GCATTAAGACTCTTGTAGATCACTTAAGAAAGAAAGTCCGACGATATGTAAGCAATTAAT 2196
Db 670 -----ValLeuAspThrLeuLysGln----- 676
QY 2197 ACAAGACAGAGATTTGAACATCCATTAATGCAACAAACCCGTGTGTTAAACCTCGT 2256
Db 677 -----CysGlnAspThrHis----- 681
QY 2257 GAAGGACGACCAACCACTTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2316

Db 681 ----- 681
QY 2317 TACGAGAACACGAAATCGCTGCTTCATAAATTGAAGAAAGCACAGAGGTATA 2376
Db 681 ----- 681
QY 2377 TATAAACGTGGGGGTAGAGAAAGACTTCACAGACAAATTATGTAGATTAATGAAAA 2436
Db 682 ----- 684
QY 2437 CATTTAATGTAATCTTGTTTTCAAATGACATGTGATGGCAAGGCACAGGTGAT 2496
Db 684 ----- 684
QY 2497 GGTATACAAACAGATTTCGTAGAACCTGAATGGGAAGTGATCCGAAACATGCGT 2556
Db 685 ----- 688
QY 2557 AAGATCAGCAGATGTTATTTATGCTCCTAGAACAGCATATATGATCATCCAAATTTG 2616
SerLeuAspValLeuLeu
689 SerLeuAspValLeuLeu
QY 2617 GAACATTACAAACGATGATCACCCTAATGTAATTTGTGATGATTTAGTTAAT 2676
Db 695 ----- 695
QY 2677 AATTCCTTTTGGGGATGTTCTTATCAGCAAAATATGACAAACAGATATAGCA 2736
Db 695 ----- 695
QY 2737 ATGTATAAAGAAAGATTAACCTAAAGGCCCAAGAGTAAGTACCCAAACACACCG 2796
LysGluLeuLeuLeuLeu
QY 2797 ACAATATCTGTGAGCTATAGCTTACAGTTTGCAGATATAGTATATTAATTCAGAGA 2856
Db 700 ----- 700
QY 2857 AGAGATCTCTGGAGAAAGACGGTGACATGTTAAGCTGCAGACATTTTGAAACTGT 2916
Db 700 ----- 700
QY 2917 TTTGGTAATATACATAAGTCCTCAAGCAAGAAATGATTAATATATGATGATGCC 2976
Db 700 ----- 700
QY 2977 CCCAAATATTTAAATTTGAGGAAATTTGGGGAAGCTAATAGCCAGCAAGTATGGGA 3036
701 ----- 708
GCCATGAAATGTGATATAAATATTTGAAGGATTAATCGGACACCAATCAGCAAGT 3096
709 ----- 718
AGTTATTCGGGATATAGTATCATACACCATTTGATGATTTATTCACCAAAAAATTTAGA 3156
719 GluTyrCys----- 721
3157 TGGATGCCGAAATGGCAGAAATGCTGCAGAGTGACAGAAAAGGATGATTAAGTTG 3216
722 ----- 732
3217 AAGGAAAGTGTAGAGAGTGAAGATTAAGATTAATGCTGACGAAAGAGAGT 3276
733 GluGlnAspAspGluGluVal----- 739
3277 GGATACAGGTGTACGAGAGTCACAGAGCTTGAATGATTAATGATTAATAGATTA 3336
739 ----- 739
3337 TGGAAAGAACATGATATATATATCAGATTAATACAAAGATTTACATGAACAAGCAAA 3396
|||||

Db 740 ----- 744
QY 3397 ATGCTGTTAGTAATAGTGTATTAAGCTTCCAGTACTGCCAAATCATATATAGACAG 3456
Db 745 ----- 754
QY 3457 AATGTTATGAAATTTTGTGCGAATTTATCCAAACAAATGTGTGCAAAAGTAAATG 3516
Db 755 ----- 758
QY 3517 GGTACTAGTATGAAAGTGTGCTCATTTGCTACTAACCACGATATGAAATGTTGGAGCA 3576
Db 759 GlyThrLysSerLeuSerCysValLysGlyAlaCysAlaLeuValLysGlyVal-----Leu 777
QY 3577 TATCTCATGATACAGCAAAATTTGATGATTCAGTACCAAAATGATGTTTGTGATGA 3636
Db 778 GlnGlnLysSerAsnLysSerLeuAspAsnCysAlaAlaLysAsnArg----- 793
QY 3637 AAAAGTATGTTAAGGTAAAGTAACGAAAAATATGCTTTAGAGATTAACCAACAGACATGAT 3696
Db 793 ----- 793
QY 3697 GGTGCGTGTGTTGTAAGTGGATGCAAAACGCAAGGTACAGATTAACGAAAAA 3756
Db 793 ----- 793
QY 3757 AAGCGAAGAAAAAGATACGAAATGTAACAGATGATGATTAATTAAGAAAGAT 3816
Db 793 ----- 793
QY 3817 GGAAGAAACAGTACAGATTTGTCATCCAAAGAAATGTAATGATATCCCATGG 3876
794 ----- 798
3877 CAATCGGAAATATTAATTTAGTGAA-----GACCCGTGTGTATGCCCCCTAGAGA 3933
Db 799 GlnCysAspLysAsnThrPheValAspGlyAsnGluValLysMetProPheArg 818
QY 3934 CAAAGTTATGCTATCTTCTGGCAATGATTAATGAATTAATTAATTAATCAATCAAA 3993
LysSerLeuLysLeuAsnLeuThrLeuGlnGluGlnThrLys-----AsnLys 835
QY 3994 GTTAATTTAAAGAGCTTCATCAATCTGCAGCAAGCAAAATCTCTCATGATAT 4053
Db 836 TyrGlnLeuArgGlnAlaPheLeuLysCysAlaAlaLysLeuThrAsnLeuLeuTyrAsp 855
QY 4054 TATTAATTAAGTAAAGTATGTTGAGGAAATGAACTGATTAAGATTAAGAAAGCAAA 4113
856 LysTyrLysAsnAspLysAsnGlnAlaGluGlnLeu-----LeuLysLysGlyLys 872
QY 4114 ATTCCTCCCGCATTTTGAAGTCCATGTTCTACACATTTGAGATTAATAGATTTTATTA 4173
Db 873 IleProGlnAspPheMetArgIleMetPheTyrThrPheLysAspPheArgAspPheCys 892
QY 4174 TTGGAAACAGATATATCAAAAGTGCATGTGAGGAAAGTAACTAAAGCAATATGAT 4233
Db 893 LeuGlnAsnAspMetGlyLys-----AspValAspLysValLysLysAsnIleAsn 909
QY 4234 TCTCTTTCAAAATGCTGACCAAAATCTCTAATGGAAGAAACACGCCAAGATGTGG 4293
Db 910 LysValPheAsnAsnSerLysArgGlyPheLysLysIleAspProGlnAsnThrPyr 929
QY 4294 ACAGAACATATGATGATGATGAGAGAGCTATGCTATGCTACATGTAATAAATTTGGGCA 4353
Db 930 AsnGlnAsnGlyProGlnIleTyrAsnGlyMetLeuCysAlaLeuIleHisAlaAspThr 949
QY 4354 AAA-----AAAGATTTTACCGAAAACTACGGTTTCAACAACGTCAAA 4398
Db 950 LysAspSerLeuLysAsnLysAsp-----AsnTyrLysTyrGluLysValThr 965
QY 4399 TTTAGTGCAGAA-----AGCACCACTTTGGAGCAATTTGCCAAACGACC 4443
Db 966 IleLeuAlaLysArgAspGlySerAsnGlyMetThrLeuSerGlnPheAlaLysPro 985

QY 4444 CAGTTTTCAGATGCTAACCGAATGGTATGCTATTTGCTATACAGCAAAATAT 4503
Db 986 LysrheleuArgrPheValGluTrpTyrAspAspTyrCysLysGluArgGlnLysTyr 1005
QY 4504 TTGAAGATGTGCAGGAAAAATGTAAGTCAATGAC-----CAATTGAAGTGTATCA 4557
Db 1006 LeuThrGluValAlaSerThrCysLysSerIleAspGluLysGlnLeuLysCysAspArg 1025
QY 4558 GAATGTAATAGAAATGCGAGGACTAGCTTAATATATATGAAAAAATA-----GAGTGC 4614
Db 1026 GlyCysAsnAsnLysCysAspGluTyrLysLysTyrMetArgLysLysGluLutrp 1045
QY 4615 ATTCCACAGATATAATATTCAAGATGAACGGACAAAAAAGATTGCTATAGCAACAC 4674
Db 1046 AsnLeuGlnAspLysTyrTyrLysAspLysArgLysGlnLysGluIleAspLysGluPro 1065
QY 4675 ATTTGGTATATGTTACAGCTTACTGAGACAAATGCAACAGATTACTTGAACAGAAA 4734
Db 1066 IleGluLysIleValLysAspTyrValLeuAlaAsnAlaLysGluLutrpLysLys 1085
QY 4735 TTTTACTGCTAGTTGTGCTAGTAACTGGAAGTGGC-----TGTGTGTA 4779
Db 1086 PheThrAlaSerCysValThrSerSerGlyLysAlaGlnAsnSerAlaThrGluVal 1105
QY 4780 CAAGAATAATATACATTTAGAAAAACAGCTTACTATGATGCGCAACATTTGGG 4839
Db 1106 LysLysAsnIleGluLeuLeuSerGluGluGlnTyrTyrAspAlaAspGlnTyrCysGly 1125
QY 4840 TGCACAAATTTATGAAATGACGACAAATATACTACATATCGTAAAGTAAGTGC 4899
Db 1126 CysThrLysPheLeu-----HisAspAspLysTyrSerLysIleSerGluYrSerAsn 1144
QY 4900 AAGAGTTAGTAAAGGACCAACACAGCTGCTATTAACTGGCAAAACAAAGCTCTAAT 4959
Db 1145 CysGlyLeuAsnSerAspAlaLysLysAsnLysIleLysTyrArgAsnSerAspGluLys 1164
QY 4960 AACTCAATAATCTGTAAGAA-----TGACTGAAGATGCTCTTTCTCTGTCGA 5013
Db 1165 AspTyrAlaPheLeuLysLysArgAsnLeuSerGlyAspValPhePheProSerArgArg 1184
QY 5014 CTACGTATATGTTTTCATGATGATGATGCAATATATACAGATCCAGATTAAGATGA 5073
Db 1185 LeuArgLysCysPheHisAlaLeuAspGlyAsnTyrThrAspProGluValLysAspGlu 1204
QY 5074 AATGGGTTGCGAAAAAGATGATGCAAGTGGCGGCAAGGATGACAAATTTGGTCAA 5133
Db 1205 AsnGlyLeuArgLysArgLeuMetGluValAlaAlaThrGluGluTyrAsnLeuSerGln 1224
QY 5134 TACTCAAAAGAAAAAAGAAAAAGAAAAATTAACAGTGGATGCGCACAAATATCT 5193
Db 1225 TyrTyrLysGluLysGluLysGluLutrpThrGluAlaThrGluGluAlaHisLysTyrSer 1244
QY 5194 TATGAGTCCCGCTTGTAGTGCCTATGAATATAGTTTATGATTTAAGATTAAT 5253
Db 1245 TyrGluValGlnProCysSerAlaMetLysTyrSerPheTyrAspLeuArgAspIleIle 1264
QY 5254 CTAGCTATTTGATTAATTTGCAAGATGAACAAAAACAGCGGAAAAATTTGAAGAAATA 5313
Db 1265 LeuGlyTyrAspAsnLeuGlnAspAsnSerThrThrThrGluLysAsnLeuLysLysIle 1284
QY 5314 TTTAACAAAAATGAAACATCACTTGGCAAGAGTATAGTACTACAGAAATCCCGGT 5373
Db 1285 Phe-----LysSerGlnSerAsnGluGly 1292
QY 5374 AGTACTGCGGCAAAATTTTGTGACGAAATAAAGAAATGTGTGGAACGCAATGATA 5433
Db 1293 SerGlnGlyArgGlnThrPheTyrAsnAsnAsnLysGlyValTyrGluAlaMetLys 1312
QY 5434 TGCAGTACAAACCTGTAGGATGATGAGAAATAGTGAATAGTGAAGAAGAGATGAA 5493
Db 1313 CysGlyTyrLysHisGlyArgAspAspGlyAsnSerAlaAsnSerAlaSerSerAspGln 1332
QY 5494 GATCTAAAAAATGTGCTTGTACCTTCAGATGATGATGATATATGAGGAAAAATCGC 5553
Db 1333 AspLeuLysLysCysAspSerValProSerAspAspTyrProMetLysLysAspArg 1352
QY 5554 GATGAAGTACACCGATACAGTTTCTTCGATGCTTGGCCGATGGGGATTAACATTTTTC 5613
Db 1353 AspGluGlyThrAlaLysGlnPheLeuArgTyrPheAlaGluTyrGluLysAspPheCys 1372
QY 5614 AAACATTAAGAAAAAGAAATGGAATAATGTAGAGGCGCTGAATGATATCTGTGGT 5673
Db 1373 LysHisLysGluLysGluLutrpLeuGlnLysLeuValGluAlaCysAsnAspTyrLysGly 1392
QY 5674 GATTAAGATTAAGAAAAAGAAATGTACAGTGGCTGTACAAATATTAATTTAT 5733
Db 1393 AspAspGlnAspLysLysLysCysThrAspAlaCysThrGlnTyrLysGlnPheIle 1412
QY 5734 AGTGAATGGAACCCACAGATGATGAACAAATCAAAAAATATGAGAAATTAACACAA 5793
Db 1413 SerGlyTyrLysProGlnTyrGluLysGlnTyrLysLysTyrGluLysHisLysAspLys 1432
QY 5794 ATATATCCGACATCCTGTGGCAAAAGATGCAGAGACGCTGCGCAATATTTAGACAA 5853
Db 1433 IleTyrSerArgHisProMetValLysGluAlaLysAspAlaGlnGluTyrLeuAspLys 1452
QY 5854 CAATTAATAAATAATTTGTGAATAAAGTGAAGATTGTGAATATATATGATAAGAT 5913
Db 1453 GlnLeuGlnLysSerCys-----AsnSerGluLysCysAsp-----CysMetAsnLys 1469
QY 5914 GTGTCCACACAGCGATTAAGTATGATGATGTAATAGTCAAAATATGCGCGCATTTAGAGAT 5973
Db 1470 LysSerThr-----SerAsnGlyAsnAsnMetProAlaSerLeuAspGlu 1484
QY 5974 GAACCAAAAGAAAGTTGAAGAAAGTAAATGTAAGTCAAGTCCACAGGTCACACGTGA 6033
Db 1485 ThrProSerThrTyrLysAspArgCysGluLysGlnProProProProPro----- 1502
QY 6034 CGAAGGAAACACCGTCACACAGGGTATCACTGATATCAAAAGCAGCGCATCGAAAA 6093
Db 1503 -----ProAlaProAlaArgProProAlaAlaArgLysSerGlyAsnAspTyr 1519
QY 6094 GAAGCAAAACA-----GGCGCGCTACAAAA 6120
Db 1520 ArgGlyArgSerGluProGluLysAspLysProLeuProLeuProLeuProProThr 1539
QY 6121 CAGCCGAAAAA----- 6132
Db 1540 ProProLysProGluLysAspGlyAlaGluArgLysIleLeuSerThrProAlaGlnGly 1559
QY 6133 -----GTGAAAATCTTAACAAGAAATGCGACACAAACACGACCGCAGCAGCA 6186
Db 1560 ThrIleGluGluAspGluGluLysAspAspGluGluGluGluLysAlaGluAlaAlaSer 1579
QY 6187 CAACAAACACGAAACGAACATCAACAGCAACAAACAAACAAATCTGACGTGGCACAATG 6246
Db 1580 GlnLutrpThrGluSerLys-----GluValValGluGlnLysGluAspThrThrGluLys 1597
QY 6247 GTAAAGGCCATCTTTGCAATTAACAGATAGCAGGCGGTGAATA-----GAGGGTGAAT 6303
Db 1598 ValCysLysIleValAlaAsnIleLeuThrGlyLysLysLeuAsnAspAspAlaCysAsn 1617
QY 6304 CAAAAACGTATGAGCAATCTCTAA-----TGGCGTGTATGTA----- 6345
Db 1618 GlnLys-----TyrGly-----TyrProGlnArgHisTyrPduTyrLysCysValThrProThr 1635
QY 6345 ----- 6345
Db 1636 ThrSerSerSerThrSerGluArgGlyAlaSerArgAsnLysArgAsnLeuAspSer 1655
QY 6346 GGTAACTTAAGAAATAAGAAATGCAATGTATGCTCTTGAAGAAAAAATATATGT 6405
Db 1656 ThrLysSerSerAspLysAsnGlySerIleCysIleProProAlaArgArgArgLysLeuTyr 1675
QY 6406 ATAAATTAATTAACA-----TATTTAATTAATGAACACTGA----- 6441

```

Db 1676 IleLysLysIleGlnGlnIuTrpAlaSerGlyAsnThrGlnIaGlyAsnThrGlySer 1695
QY 6442 -----AATAGCGTGCACATGATATATA----- 6465
Db 1696 GlyAspSerThrGlnLysAlaSerSerAsnProGlnGlyAsnGlyValSerThrSerProGln 1715
QY 6466 -----GAGGCTTTTATTAATGTGCAGCAATAGCAATGCAATTTTGTGGTGA 6513
Db 1716 ValAlaLeuLeuHisAlaPheValGlySerAlaValGlnIuThrPheLeuTrpAsp 1735
QY 6514 AATATAT----- 6528
Db 1736 ArgTyrLysLysGlnLysGlnIleGlnLysGlnGlnIuThrGlyLeuValAla 1755
QY 6529 AATCGTCGACAGAA-----AATGAAATTCGAAATGGAACAAATTCGA 6570
Db 1756 SerGlnIuThrSerGlnAspThrGlnHisProGlnAsnLysLeuGlnSerGlyThrIlePro 1775
QY 6571 GATGAATTTAAAGAAATATGATATATACATATGCTGATTAATAAGATATGTTTGTGA 6630
Db 1776 LeuAspPheLeuArgGlnMetPheTyrThrLeuGlyAspTyrArgAspIleCysValGly 1795
QY 6631 ACAGATATTTCTATGATATAAAAAATTAATACGTAACTAACAATAGCTGAACACATTCG 6690
Db 1796 -----LysThrProAspGlyIleAspThrValSerIaSerAspLysAspThrMet 1812
QY 6691 AATGAAATATTAATAAGAAAAACAGAT-----AAAAAAAGAT----- 6729
Db 1813 AspLysIleGlnIaLysIleGlnGlnIleLeuProLysLysAspThrProProSerVal 1832
QY 6730 GAAGAAATTAACGTAATAATATTGGGAGAAAAATTAATAATTTATTTGGAGAAATGATA 6789
Db 1833 LysThrProGlnIuThrTrpAsnIleAlaGlnSerIleTrpAsnGlyMetIle 1852
QY 6790 TATGATTAATCTATGATCTCACAGACGAAAGAAAAAATAATATA----- 6840
Db 1853 TyrAlaLeuThrTyrLys-----ThrAspThrProSerGlyLysProLysGlnIlePro 1871
QY 6841 -----GATTAATTCACGATAC 6855
Db 1872 GlnValLysThrLysLeuPheAspGlnLysGlyThrProGlnSerAsnLysTyrGlnIuThr 1891
QY 6856 AATGACATGACCAACTG----- 6873
Db 1892 LysThrVal-----LysLeuGlnGlnIuGlnIuThrSerGlyAlaLysProLysSerThrGln 1910
QY 6874 -----ACGCGCTTCC-----CTTGAAAGCTTTGTAAAAAGCCCAATTT 6912
Db 1911 SerSerSerProSerGlnGlnAsnThrProLeuThrAspPheIleSerArgProTyr 1930
QY 6913 TTGAGATGTTTCACAGATAGCGCAGAGAAATTTTGTAAAGAGAGAAAGATGTTGA 6972
Db 1931 PheArgTyrLeuGlnIuTrpGlnIuAsnPheCysLysGlnIuArgLysLysArgLeuGly 1950
QY 6973 AAATTCGAGCGCGCTGT-----AAGCAATATGAGTAAATGCT----- 7011
Db 1951 AspIleLysSerAsnLysLeuLysAspGlyAspLysGlnIuThrSerGlyAspGlnIuAla 1970
QY 7012 -----AGTAAT-----GACGGTAGACACAAAGATGT 7038
Db 1971 CysSerAsnIleAspValAsnLysAspLysIlePheAlaAspLeuGlnIuProLysCys 1990
QY 7039 GCAGAGCGCTGTGTACATATCAAAATTTTATTAAGAAAGTGAAGAACTGATATGAAGA 7098
Db 1991 AlaLysProCysSerSerTyrArgLysTrpIleLysGlyLysLysThrGlnIuGlnLys 2010
QY 7099 CAAGAGAAAAAGTTCAAAAAAGATTAAGATGSCAAAAAGATTAAG-----GATTAAT 7149
Db 2011 GlnIuLysLysLysLysGlnIuGlnIuGlnIuGlnIuGlnIuGlnIuGlnIuGlnIuGlnIu 2030
QY 7150 CCTTCTACTGAAGACATAGAGAAAGCAACATGCTCATGAAATTTTAAACATGA 7209

```

```

Db 2031 AsnGlyPheArgValLysLeuGlyThrCysThrThrAlaGlyAspPheLeu----- 2047
QY 7210 TTAAGAAATTAATGCGCATTAAGATTTGTTCTGTGATGCAAAAACCTTCTCAACAATA 7269
Db 2048 -----GlnThrLeuLysAsnGlyProCys----- 2055
QY 7270 CCAAAAAACACACACAAATCACAATCCAGATCCATATGATATGCCAGAAATCGTGAT 7329
Db 2056 -----LysSerGlnAsnGlyLysAspHisGlnIu----- 2065
QY 7330 TATGTTCTGACAAATTTAACAAAGTGTGAGTGTCTGCACTTTCAAAAAAGGATCTATG 7389
Db 2066 -----AspIuIleAsnPheSerGln---ProAspValThrPheArgProAla--- 2080
QY 7390 ATTCAACAAAAAATAATTAATGAACTTAATAATACCTATGATGTGTAAGAAAGACGA 7449
Db 2081 -----ThrAsnCysThrProCysProLysPheLysValAsnCys----- 2093
QY 7450 TATTAATTAATCTAAGACGAGAAATTAATATGATATTAATCTTGAAAGAAATTTATA 7509
Db 2093 ----- 2093
QY 7510 CTTATGAGTCTCAAAAGAAAGAAAGAAAGTAAATATGTTGACATATATATCTTGC 7569
Db 2094 -----LysAsnGlyAsnCysGlyAlaAspThrAsnGlyLysCys 2106
QY 7570 GATCTTAAGAAACCTTATGACCTGATTAATATATAGAGAAAGAAACCTTGTAATAAT 7629
Db 2107 AsnGlyLysThrPro----- 2111
QY 7630 AGAGAAAGAAATCGTTTAAAGTAGATTAATTAATGAATGTAACAAATTCAAAGTTC 7689
Db 2111 ----- 2111
QY 7690 TATCAGAGAAAAAAGATGATGTCTGCTCCAAAGAAAGAAACATATGCTTAAGAAAT 7749
Db 2111 ----- 2111
QY 7750 TTAGATGAATTAATAATTGAAGACTTAAGATAGTAAATATATCTTAATAATGTTGCT 7809
Db 2112 IleAspAlaGlnAsnPheGlnGlnIuGlnIuMetGlyGlnThrAlaLysGlnPheValMetLeuVal 2131
QY 7810 CGAATGTCACGAAATGAGAAATGACATATATAAAAAATTTCAACTCAGAGAAAGGCTGC 7869
Db 2132 SerAspLysSerThrAsnGlyPheGlnVal-----AsnAspLeuAsnGlnLys 2147
QY 7870 GCAATGAATCCAAATATGTATACATATGAATATATAGTTTGCCTGATCGGCTGACATAGT 7929
Db 2148 Ala----- 2148
QY 7930 AGAGAAACAGATATGTTACGAATTTGTTACTTACCTCCGTAGAAATTAATATATAT 7989
Db 2149 -----GlyLysAlaAsp----- 2151
QY 7990 AAGCTTTTGAATACATATATGCAAAATGAGAAATTAATAAGGTAGAAATTAATAC 8049
Db 2152 -----IlePheGlnGlyIle----- 2156
QY 8050 AACGATGTACAAACGTTCTGCTTGTGTTGGATGCTAATGAAGAAAGATTTGGAAA 8109
Db 2157 -----LysGlnAsnIleTrpSer 2162
QY 8110 GCAATG-----ACGTGCAAAAGCCACAGAA-----GATGCAAAA 8142
Db 2163 CysArgAsnValCysGlyLeuAspValCysLys-----ProGlnLysValAsnAspGlnLys 2181
QY 8143 CTTTATTTGAAGAAGAGAAATGATGATGATTTGAACGCCATTAATTAACAAGATAGCT 8202
Db 2182 ValAsnGlyLys-----GlnAsnAspGly----- 2189
QY 8203 GGAATGAAGAGACATCACTGCTGATGATTAATATACCAACGGTTTCGATGGATGACT 8262
Db 2190 -----ThrTyrIleIleGlnIleArgAlaLeuLeuArg 2200

```

Qy	8263	GAATGGCTGCATATTATTATGTAAGCAGCTATGGAAGATTTGGAAAAATTTTAAAAATCA	8322
Db	2201	ArgTTPValGIlaSnpHe	2216
Qy	8323	TGGATCACTGTAAACATCTGCACAGATGCAAGATGATTAATGATAAATTAAGTGCAA	8382
Db	2217	IleSerHisCysThrAsnSerThrGIuGIuLys	2232
Qy	8383	CAGGTAAACAGAGATGTCAGAAATATAAAAATTTTGTTCTTAATGAAATCTGTATTC	8442
Db	2233	---CysGIuLysLysCysLys	2247
Qy	8443	GATATACATCAATCAATTAATCAAGAAATGTATGAACAACCATATATATCAAAAAATCTCT	8502
Db	2248	LysGIuGIuTrpLysLysIleLysGIuHisTyr	2258
Qy	8503	ACTTATGATCATGTTCAAAATTTTGTAACAAAGCTGAAACCTTTTAAAGTGATGTTCT	8562
Db	2259	-----ValGIuGIuAsnLysSerGIu	2265
Qy	8563	GTTGAGACCTTTTCTGAATATCTTCATGAAACAAGTAAGTTGATTAATTAATTAAT	8622
Db	2266	---AspAspPheSerAspAsnLeuAsnSerPheLeuGIuTrpLeuIleThrGIuIlePro	2284
Qy	8623	CAAAATGATGGTTCCTCCAAATATACGAACATATGCTTTTGAAAGAAACACCAAAAGTTAT	8682
Db	2285	ValIalaAspValGIuGIuAsnValIleLysLeuSerAsnPheAspThrPro	2301
Qy	8683	AAAGAGCTTCAGCTTTGTACACTACCTCTCAAGATCCATTTGGANAATGTTCCTACCGAT	8742
Db	2302	-----CysGIuLysSerAlaAspAlaAsnSer	2310
Qy	8743	CAAAACAAGATGGATTAAGATTAACAAACTTTTACCTTGCTCGAAGATGATTAAT	8802
Db	2311	GIuLysLysAspGIu	2320
Qy	8803	GATATATCTTGATTAATTTGGAACGCATCTGTCTTAATAGTTGCAGATGATTAACAA	8862
Db	2321	IleAspCysMetIleAsn	2333
Qy	8863	GGCTATATGATTCCTCCAAAGACATTAATGACAAAGACATTCAGCATTAAT	8922
Db	2334	AsnCysGIuTrpGIuProSerGIuSerLysGIuCysThrTrpProTrpThrLeu	2352
Qy	8923	TATGAAAAAGGTGATTAAGAAATTTTAAAAAAAACCTTCTACTCTGCTTCAGTCAA	8982
Db	2352	-----	2352
Qy	8983	GGACAAATGTTAGGTCAAAAATTAATAATCGAAGAAGATTGTGCTTTGAGGCATGA	9042
Db	2353	-----GluAspAspGIuTrpPheAsp	2359
Qy	9043	TATATGTTATGCAGATTATTCGATATTAATTAAGAACTGATATGATGACACTTCATTA	9102
Db	2360	-----AspAspIle	2362
Qy	9103	TCTCAAAAAATTAATAAATATTTGAAACATCAATTAAGAACACCGAAATCGTAAACA	9162
Db	2363	-----GluTrpAspAsn	2366
Qy	9163	TGGTGGAAAAATATAGAGCTCAGATGATGCACCGCTATGATGTGATTAATAATTCGT	9222
Db	2367	-----ProValIalaHisProGIuIleCysGIuGIuValIleGIu	2379
Qy	9223	ACTTCAAAAGTAAACATTAATGATGAAAGATGCTGTCAATTAACCAAAAGATGAGAATTAAT	9282
Db	2380	ThrGIuTrpGIuTrpValLeuGIuGIuLysCysValIalaProValGIuGIuLysLys	2399
Qy	9283	CAGTTCTTCGTTGTTATTAATGAATGGGCAAGAAGATGTAAGAAAGAAACATGTA	9342
Db	2400	Asp-----GluLysLysLysPro	2405

QY	9343	AGTCATTCATTAATAAACAATAATGCTCCGCTTCACAAACGAAAGTAATTTTGAAGCTCAGAA	9402
Db	2406	GIUasppro-----	2408
QY	9403	TTATTAAGACACCTGGATGTCAGAAATGATTAATAATATATTAGCTTGAAATATATTG	9462
Db	2408	-----	2408
QY	9463	ATATAAATAATCAATGAAAAATCTAAATATATAATATAGCAATTTAAAAGATCAATCTTCA	9522
Db	2409	-----	2413
QY	9523	GGTAATATAGACAAATAAA-----CCATCTGAAGAAAAATGTTCACTCATATATA	9570
Db	2414	GIyAlaIleGlyProserGlyProAlaGlyProGlnGlnProserProThnSer-----	2431
QY	9571	AAATCAAAAGATTTCTCAATGCGCTTGAGACTTAATATGATTAATGAATAGTTATTCACGA	9630
Db	2431	-----	2431
QY	9631	ACAAAAATATGAAAAATATGAAATTCAGAAAGACTATAAAAAATATATCTCGTTTA	9690
Db	2431	-----	2431
QY	9691	TATTTTGTGAAGATGAAACACACAAAAATCATGTACTAGATGGAATATATAAAGAA	9750
Db	2432	-----	2439
QY	9751	GAGCAAAAGATGTGCTCCTTAAGACACTATTTCTTTACACCCCATGAGATTTCTTCTAT	9810
Db	2440	-----	2447
QY	9811	CAAGACCTTTATCTTCACACACATCGAGTAGACAAA-----	9846
Db	2448	ProAlapro-----	2465
QY	9847	---TATGATCCTAAAATATGATATATTGAAAAAGTAGATCTCTGTTGTTATTTGATCGCGC	9903
Db	2466	ProPhasppro-----	2483
QY	9904	TTAGTTTATAGCGCTTCATTCATGAGAAAAAATTCAGATCGTGGAGCTTGTTG	9963
Db	2484	LeuGlySerIleIaphhepHeMetLysLysThLysIleProValAspLeuPhe	2503
QY	9964	CGATATCTAAATATCCCGCAGAGACGATGAGATGCTTACGTTGGAATCCAAAAATAG	10023
Db	2504	SeValIleAsnThProLysGlyAspTryAspIleProThrLeuYsSerLysAsnArg	2523
QY	10024	TACATACCATATAGAGTGTGCTCATTTAAAGCAAAACATATATATATATGGAAGAGAT	10083
Db	2524	TYrIleProTyArgSerLyseryLysGlyLysThYrIleYrMetGlnGlyAsp	2543
QY	10084	ACTAGTGGAGATGAGATAATATATATGTGGACTTATCTCTCGATATTAATCTTATCC	10143
Db	2544	Set-----	2561
QY	10144	GAAGAGGATATGAGAAATGGATATTAATATATATATAGTACCGAGTATGCTCTAATAT	10203
Db	2562	GIUserGlyTyrGlnGlnLeuAspIleAsnAspIleTyValProGlySerProLysTy	2581
QY	10204	AAAACATTCATGAGTAGTACTATACCAACCATCAAAAAGG-----GATATACCA---AGT	10254
Db	2582	LysThrIleIleGlnValValLeuGlnProserLysSerAsnGlyAsnIleProThnSer	2601
QY	10255	GATGATACCAACCAATGAT---ACACCAGTACGAATAGATTATTTAGTGAATAGG	10311
Db	2602	AlaGlyIuProLeuAspAspMetValProThrThrAsnThrPheThrAspIuGlnIuTrp	2621
QY	10312	AATGAACGAAACATGATTTTGTTATCTCAATATATACAAATACGAACCAATATAT---	10368
Db	2622	AsnGlnLeuLysIleAspPheIleSerGlnTyrlIleGlnSerArgLeuProMetAspVal	2641
QY	10369	---AATTCACAA---AGTGCAGATATTTCCAAATGAAATACGAACCTTAATACTTATATCT	10422

```

Db      2642  ProGlnTyrAspValSerThrGluLeuProMetAsnIleValGlnValLeu---Asp 2660
OY      10423  GATAATCTCGTGAAGAAAACCTTTATTTATATTCATTCATAGAGATTTATATGAGG 10482
Db      2661  AspGlyIleAsnGlnLysProPheIleThrSerIleHisAspArgLysLeuTyrAsnGly 2680
OY      10483  AAGAAATATGTTATATATATATATGATGACTAATATATATATATATATATATATAT 10542
Db      2681  GlnGlnIleSerTyrAsnIleAsnMet-----ValAsnAsnAspIleProMetSer 2697
OY      10543  GCTAGAATGATCTTATATAGAGTATA 10569
Db      2698  AspLysAsnGlnTyrTyrThrGlyIle 2706

RESULT 2
T28625
variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
Accession: T28625
Su, X.Z.; Heatwole, V.M.; Werthelmer, S.P.; Gulnet, F.; Herfelde, J.A.; Peterson, D.S.
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
A:Reference number: Z20487; MIMD:95330813; PMID:7606788
A:Accession: T28625
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3006 <SD>
A:Cross-references: EMBL:L40609; NID:g886376; PID:g886377; PIDN:AAA75397.1
C:Genetics:
A:Gene: var-3
A:Introns: 2597/3

Alignment Scores:
Pred. No.: 4,28e-202 Length: 3006
Score: 4260.00 Matches: 1143
Percent Similarity: 43.848 Conservative: 483
Best Local Similarity: 30.828 Mismatches: 1005
Query Match: 21,904 Indels: 1079
Db: Gaps: 134

US-10-087-013-1 (1-10628) x T28625 (1-3006)
OY      91  AGTCACAAAGTGCAGAAATGTTTGAAGCTTATGCCAAAAATATAGACAT---CCA 147
Db      8  ThrAsnLysThrAlaLeuThrTyrAlaAsnAspLeuGlnIleGlyLysGlnValGlnGlnLysAla 27
OY      148  TCAAAATATGCCAAAGACATGTCATTCGTTGAAGGGGATTTGACGAAGCAGAAATTT 207
Db      28  ThrGlnAspAlaLeuThrTyrAlaAsnAspLeuGlnIleGlyLysGlnValGlnGlnLysPhe 47
OY      208  CGTGTGTCCTCTACAGCCAGTAAATAGCATATATATATATATATATATATATATAT 267
Db      48  HisGly-----ValProIleAspValLysAsn-----ProCysAsp 59
OY      268  TTAGATCATAGAGACATCTAATTTACGGTATGATGATGATGATGATGATGATGATG 327
Db      60  LeuAsnTyrGlnIleHisThrAsnVal---ValLysGlnTyrGlnLysGlnLysGlnLys 78
OY      328  CATGTGATAGAGACAAACCATTTGATGAGAGATGAGACATCTGAATGCGAAATATA 387
Db      79  ArgGlyTyrGlnIleLysAspPheSerAspValLeuSerGlyGlnSalAlaLysAsnThr 98
OY      388  CGTAATATATTAAGAAAATATGATGCTATAGCTGTGCGCACCTACAGAGCAGATATG 447
Db      99  IleLysAspSerValThrAsnSerValGlyAlaCysAlaProPheArgGlnLeuHisVal 118
OY      448  TGTGATTAAGAACTTGAGAGCTTAATGATATATATATATATATATATATATATAT 507
Db      119  CysAspArgLysLeuGlnLysLeuIleLysThrAspGlnAsnThrSerThrHisAspIleLeu 138
OY      508  GGAATGTACTACTATACAGCAAAATACGAAGTGCAATCTGTATATATATATATAT 567

```

```

Db      139  ValAspValLeuLeuThrAlaLysHisGlnGlnSerLeuValLysHisTyrLysGln 158
OY      568  -----AAGGAACCTCAGACGCTTGACCTCTCTGCACGAACTTTTGCAGAT 615
Db      159  TyrIleLysLysAsnArgAsnPheAsnIleCysThrValLeuAlaArgSerPheAlaAsp 178
OY      616  ATAGTGATATTTGTAAAGCAATAGATATGTT-----AAACCAAT 657
Db      179  IleGlyAspIleValArgGlyLysAspLeuTyrLeuGlnTyrAspAspGlnGlnLysGln 198
OY      658  GTCCATGACCAAGTAGAAGAGGGGTCCGAGAGCTTTTCAAGAAATACATGATGATG 717
Db      199  LysArgLysGlnLeuGlnLysAsnLeuLysPhePheGlnLysIleHisAspAspVal 218
OY      718  -----GAAGATGACGTAATAATATGATTCAT 744
Db      219  MetLysThrSerGlyArgThrAsnGlnLysLysSerAlaGlnValGlnLysArgTyrAsn 238
OY      745  CCGATGATCTGGAATATTTATTAATTAAGAGACATGGTGAATGTGAATAGAAAT 804
Db      239  ---AspAlaThrGlnLysAsnTyrTyrLysLeuArgGlnAspTyrPheAsnAlaAsnArgAsp 257
OY      805  AAGATGGAAGACCTATACATGTCATCATATATATATATATATATATATATATAT 864
Db      258  GlnValTyrLysAlaIleThrCysAspAlaAlaAspAsnAspGlnTyrPheGlnAsnSer 277
OY      865  GAAAGTATACACCATATTTTCAATCTTAATTCGGCCATTAACAGGAAGAGTTCT 924
Db      278  SerAspGlyLeuTyrValPheSerAsnGlnCysGlyArgAsnGlnGlnLysValPro 297
OY      925  ACCAATTTAGATATATGCTCCCTCAATATTTAGCTGTTCACGATGCGGAGAGATTT 984
Db      298  ThrAsnLeuAspTyrValProGlnHisLeuArgTyrPheAspGlnTyrPheAlaAspPhe 317
OY      985  TGGCGAAAGAAATATATATATTAATTAAGCAAGACCTCCCTGCTATGACAAAGAA 1044
Db      318  CysArgLysArgAsnIleThrLeuLysSerAlaLysGlnLysCysGlnGlnGlnLysAsn 337
OY      1045  CGCTTATATGATGATCATTAATGACATGATGTCGACAACTATTTGAAAAAAGTATT 1104
Db      338  ThrLysTyrCysSerLeuAsnGlnTyrAspCysThrLysValPheGlnLysLysAspSer 357
OY      1105  TTGCATTTGATATTAAGTGTACTGACTGTTCGACTAATATGCAAACTTTTGAAGTTGG 1164
Db      358  CysSerSerAspGlyAsnGlnThrAlaCysSerAsnIleCysIleAlaTyrAspAlaTyr 377
OY      1165  TTAGGCAATCAACAGAGACATTTAAAAACAAAAACAAATATGAAAAAGAAATCA 1224
Db      378  LeuArgAsnGlnArgAsnGlnPheGlnLysGlnLysIleLysTyrThrLysGlnIleGln 397
OY      1225  TCAATATTTGCAACGATCAAAATTTGCAATATATATATATATATATATATATAT 1284
Db      398  LysTyrLysSerSerSerAspLysSerAsnSerAsnIleSerAsnLysTyrTyrAsnGln 417
OY      1285  TTTTATGAAAAAAGCTTAAGAGAAACCAATATGCAATATGACATCTTTTAAATTTACT 1344
Db      418  PheTyrGlnAsnPheGlnLysLysGlnTyrGlnIleThrLeuGlnAsnPheLeuLysLeu 437
OY      1345  AATGACGAAGATATTTTAAGAGGATTTACACGAGAAAGATATATCTTTTACTTAAC 1404
Db      438  AsnLysGlnLysTyrCysGlnGlnLysIleGlnIleGlnIleValIleAspPhe----- 455
OY      1405  AGTGCATGCAAAAGGATTTTATCGTTAGAAATTTGCGCAAGTGTCCGCGATGC 1464
Db      456  AsnLysAspGlnLysPheMetValPheHisArgSerGlnTyrCysGlnProCysProAspCys 475
OY      1465  GGGGTCAATGTGATGCTAATAATACACACAAATACATATGATCGTGAACGT--- 1521
Db      476  ValValGlnCysLysGlnGlnLysCysThrGlnAspLysLysAsnAspLysCysArgSer 495
OY      1522  -----GTAATTAATGAAGACATATTAACCTCATGCGGTGGAACCT 1563

```



```

Db      496  LysIleIleIleYbAlSyleuendInserGlu-----Pro 507
QY      1564  ACTAAATATCAGTCCTCTTATATGCTAATAGCAACAGCGTATTTACACAAAATTAGAA 162
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      508  ThrGluIleHisValIleuSnsrSerpaspysgInciAspIleThrLysIleuGlu 527
QY      1624  AATTTTGTAAACGCTCAACTATATTAACAGCTTAAAAATATATCAAAATGGGAATGCTAT 168
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      528  ValPheCysSerSerThrThrAsnIlytIuclYrAsnValGlnIlyStrpLysCysYr 547
QY      1684  TATACGATGAAATATTAATATAGATGTAACCTGGACAAACAAATCTGAAATCAATATGAT 174
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      548  AsnLysAsnSerAspLysAsnSncysgIuMetAsnIleSerSerYrLysAspSerThr 567
QY      1744  AATCCTAAGATATATATTCATTAATTTTGTAAATTAATGCGTTACATATTTATTAAGG 180
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      568  AspAlaAsnValMetLeuSerValGluCysPheHisSerTrpAlaLysAsnLeuIle 587
QY      1804  GATACTATTATTAAGTGAAATGACAAACTTAAACTGTATTAATATATCA---ACCAGCAT 186
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      588  AspThrIleLysTrpGluHisGlnLeuLysSncysIleAsnSnrThrAsnValThrYr 607
QY      1861  TCGATTGATGAATGTAAACAGAAATTCCTTATGTTTATGACAGATGCGTTAAACAAAAGAA 1920
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      608  CysGluSerLysCysIleLysAsnSncysGluCysTrpGluIlyStrpLileysArgLysGlu 627
QY      1921  GAAGATGAATAGATGAATAAACAACTGTCACAAAACAAAAGAAATATACGAATCGTAT 1980
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      628  HisGluTrpGluLysValLysAsnValPheGluYAsnSnsrNArgMetSerYrIleYr 647
QY      1981  TATATCTATATTAATTAATCTTTTGAAGGTTATTTTAAACTTATGCTAAATCTGAC 2040
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      648  TyrAsnAsnLeuSerArgValPheAspSerPheLeuPheGluValMetPheAlaLeuasp 667
QY      2041  AAGATGAA---GCAAAATGGAAGAACTTATGGAAGAAATATTAAGAAACAAAATATGAG 2097
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      668  GlnAspGluLysGlyLysTrpAspGlnPheThrGlnAspLeuLys---LysLysPheGlu 686
QY      2098  TTTCACATTTGGAAATATATAGGACATATTTAGACATGCAATAGCAACTGTGTTAGAT 2157
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      687  ProSerLysThrAsnThrProThrGlyLysSerGlnAspAlaIleGluPheLeuLeuasp 706
QY      2158  CACTTAAAGAAAGAACTGCGACGATATGTAAGACAAATATTAACAAAGCAAGCTGAACA 2217
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      707  HisIleuLysAspAsnAlaIleuThrCysArgAspAsnSnrSerSncInserCysAspAl 726
QY      2218  TCCCATTAATGCACACAAACCCGCTGTTTAAACCTCGTGAGGACGCAACCCACTAAA 22717
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        SerLysLysValLysThrAsnProCysGlyLysAsnProSerAlaSerAsnSnrLeuVal 746
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        AATATTAAGAAAGAAATAGCAATACTTTAAAGAGAGCGCATACGAGAACGACGAATCGT 2337
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        ArgValIleLysArgLeuAlaGluMetMetGlnArgTyrAlaArgLysGlnLeuGluLysArg 766
QY      2338  GCGTCCTCAT---AAATGAAAGAGAAAGGACGACACAGATATATTAACAGTGGGGGTAGG 2394
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        GlyGlyGluIleAsnLeuLysGlyAspAlaThrLysGlyThrTyrArgGlnGlyLysPro 786
QY      2395  AGAAGAGCTTCAAGGACGACATTTATAGCAATATGATTAACATCTT-----AAT 2445
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        AlaAspLysPheLys---AsnValCysSerIleAsnGlnAsnHisThrAsnValGlnAsn 805
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        CGTATCTTGCGTTTCA---AATGACACATGATGATGCAAA---GGCACAGGTAGGT 2499
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        AsnAsnArgAlaTyrThrTyrGlnGlyProCysThrGlyLysAspCylSerAsnGlyLys 825
QY      2500  ATACAAACAGATTTGCGTAGAAGACTGAATGGGAAGTGGATCCGGAACACATCGTAAA 2559
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        Val-----ArgMetLysIleGlyThrProTrpLys---ProGlyArgGlnIleGlnMet 842
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        GATACACGAATCTTATATATGCTCTCTACAAACGCAATATAGTACATTCATTTGAA 2619
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        SerIleGlnAspIleYrMetProPheArgGlnHisMetCysThrSerAsnLeuGlu 862

```


Db 1711 -----GlycylglutathrinalaaspHisasp 1721
QY 5633 GCTGCGAATATTTAGACAAACAATTAATAAATTTGTGTAATAAAGTGAGATGT 5892
Db 1722 ThrThrIugluThrValaspGln---GluLysAlaGluLysAspLysAspGlyGly 1740
QY 5693 GAATATAGGTATGAAGAAGTGTGTCACACACGCGATTAAGTGAATAGTCAAAAT 5952
Db 1741 Glu-----ThrProGlnLysGluThr----- 1747
QY 5953 ATGCCCATCATATAGACATGAACCAAAAGATTGAAGGAAGTGAATTTGTCAAGTG 6012
Db 1748 -----GlnProLysValGlnValAsnProCysAsp----- 1757
QY 6013 CCACGAGGTCCACACGCTGTACGAGGAGAAACCCGACCGGATTCATCATATCA 6072
Db 1758 -----LysValLys 1760
QY 6073 AAAGCGAGGATCGAAGAAAGAAAGCAACGCCGCTTCAAAACGCCGAAAGAA 6132
Db 1761 ThrLeuPheThrThrThrGluThrGluThrLysGluAlaCysProThrLysThrValAsnGly 1780
QY 6133 GTGGAAATCTTAACAACAGAAATGCGACACAACACGAAACCCGACGAGACACAACA 6192
Db 1781 ArgGlu----- 1782
QY 6193 ACACGAAACGAAACATCAACACGACACACAGAAATCTGACGTGGCACAATGTGAAG 6252
Db 1782 ----- 1782
QY 6253 GCCATCTTTCGAAATTAACACAGATAGCAGGGGTGAATAGAGGTGTATCCAAAACG 6312
Db 1782 ----- 1782
QY 6313 TATGCAATATCTTAATAGCGGTGTATGTAGTAACTTAAGAAATGAATGCGC 6372
Db 1783 -----LysPheProAsnTrpLysCysIle---SerSerGlySerAspLysSerGlySer 1799
QY 6373 ATATGATGCTCTCTCTGAGGAAAGAAATATATATTAATAAATATCAATATTTAAATAT 6432
Db 1800 IleCysIleProProArgLysGluThrLysLysLysLysLysLysLysLysLysLys 1819
QY 6433 -----GAACTGAAATAAGCGTGAACATGAT---ATAAAGAGGCT 6471
Db 1820 ThrValSerSerAspGlyGluThrThrProIleThrHisAspAlaLeuArgGluAla 1839
QY 6472 TTTATTAATATGACGACATAGAACATCAATTTTGTGTTAAATAT---ATAATGAA 6528
Db 1840 PheIleGlnThrAlaAlaValGluThrPhePheLeuTrpHisArgTyrLysLysLys 1859
QY 6529 AATCTGACGACGAAATGAATTCGAAATGGAACA----- 6564
Db 1860 GluLysGluArgGlnGluGlnLeuGlnAsnGlyThrPheLeuLeuProProAlaGlnLys 1879
QY 6565 -----ATTCCA 6570
Db 1880 ValSerProGluAspAsnProGluHisProGlnLysLysLysGluGlyLysLysLys 1899
QY 6571 GATGAATTTAAAGAAATATGATATATATATATATATATATATATATATATATAT 6630
Db 1900 GluGluPheLysArgGlnMetPheThrLeuGlyAspTyrArgAspLeuGlyValGly 1919
QY 6631 ACT-----GATATTT-----TCATGAATGAA 6651
Db 1920 ValLysAspAspValAlaGlnAlaLeuGlnLysLysSerGlyAspAsnLysSerGlyAspLys 1939
QY 6652 AAAATTTATATCTGTAACAATAATAGTGAACAACATTTCTCATGAATAATAT--- 6702
Db 1940 AsnIleLysAspIleSerGluLysIleLysSerValIleGluLysSerGlyGluGlnThr 1959
QY 6703 -----AAGAAAGAAACGAGATTAAGAAAGAAAGATGAAGATTAACGTAATAATTT 6750
Db 1960 ProProGlyProLysProGlyGlnThrThrThrLysProGluGln-----Trp 1975

QY 6751 TGGAGAAATATTAATAATTTATTTGGAAGCAATGATATATGATTAATCTATCATCTC 6810
Db 1976 TrpGlnLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1994
QY 6811 ACAGAC-----GAAACGAAAGAA 6831
Db 1995 ThrAspTrpArgGlnValAspAspGlnValLysGlyGlnLeuPheGlnAsnGlyLysAsn 2014
QY 6832 AAAATTAGATATAT---TACAGATACAAATGACATGACCAATGACGCTTGC----- 6882
Db 2015 ThrProLysAsnSerGlnTyrGlnTyrLysAsnValThrIleSerSerValSerSerLysn 2034
QY 6883 -----CTGAGAGGTTGTAAAAGCCCAATTTTGGA 6918
Db 2035 GlyGlyProIleGlyAsnIleLysLeuGlnGlnPheAlaSerArgProThrPheLeuArg 2054
QY 6919 TGGTTCACAGAAATGGGACAGAAATTTGTATTAAGAGAGAGACAGTGTAAATTTG 6978
Db 2055 TrpLeuGluGluTrpGlyGluGluPheCysArgLysGlnTyrHisLysLeuGluArgIle 2074
QY 6979 GAGCGGGCTGT---AAGCAATATAGTGTAAATGCTAATGACGCT----- 7023
Db 2075 LysGluGluCysHisLysAspLysAsnArgAsnCysAspAspAspGlyPheGluCysLys 2094
QY 7024 -----AGACACAGAAATGTGACAG 7044
Db 2095 GluMetCysProCysLysAspGlySerPheGluThrLeuLysCysProSerCysAlaLys 2114
QY 7045 GCGTGTGATACATATCAAAATTTATTAAGAGAGAGAAATGATATGAAGAGCAAGA 7104
Db 2115 SerCysLysSerTyrLysLysTrpLysLeuArgLysLysAspGluPheThrLysGlnLys 2134
QY 7105 -----GAAAGTTCAAAAGATTAAGATGCGCAAAAG-----TATAGCATAT 7149
Db 2135 GlyAlaTyrGluLysGlnLysLysAspAlaGluLysAsnAsnAspTyrLysGluPhe 2154
QY 7150 CTTCTACTGAAGAGACATAGAGAGACATAGTGTCTATGAAATTTATTAACATGAAA 7209
Db 2155 SerLysThrLeuArgAsnTyrAsnAspAlaAla-----PheLeuAsn---Ser 2170
QY 7210 TTAAGAATATATGTGCAATTAAGATTTCTTGTATGCAAAACCTTTCACAACTA 7269
Db 2171 LeuLys-----AsnGlyProCysSer----- 2177
QY 7270 CCAAAACACACAAACAATCAATCATCCGATGCTAATGATATGCCAGATGCTGAT 7329
Db 2178 -----LysAsnAspAspAspSerValGlnAspGluIleLys 2189
QY 7330 TATGTTCTGAGAA-----TTTAACAGTGTAGTGTCTGAACTTCAAAAAG 7380
Db 2190 PheAspAspGluArgLysThrPheGlyHisGluThrTyrCysLysProCysSerLys--- 2208
QY 7381 GGATCTATGATTCATACAAAAAAATTTACTGAACCTTAATACCTATGATGTGTAGAG 7440
Db 2208 ----- 2208
QY 7441 AAAGACACATATATATATTAAGAGAGAGAGAAATATATGATATATACCTGAAGAA 7500
Db 2209 ----- 7500
QY 7501 AAATTTATACCTTTGAGTCTACAAAGAAAGAAAGTAATAATAGTTGAGATATAT 7560
Db 2214 Lys-----LysValLysCys 2213
QY 7561 AATCTTGCATCTTAAGAACTTATGACCGTATATATATATATATATATATATATAT 7620
Db 2217 AsnHisCysAspAsnSerLysPro-----AsnAsp 2226
QY 7621 TGTGAAATTAGAGAGAAATCGTTTAAGGTAGATATATGATGAATGTAACAAAT 7680
Db 2227 Cys----- 2227

QY 7681 TCAAGTCTATCAGAGAGAAAAAGAGTATGTCTACCTCCAGAGAGACATATGTGC 7740
Db 2227 ----- 2227
QY 7741 TTAAGATTTAGATGAATTAATTAAGAGACTTAAGATAGTATATCTCTTAA 7800
Db 2228 ---ATGAsnIleAsnSerIleSerAlaGlu----- 2236
QY 7801 ATGGTTCGTCAACGTCACGCAAAATGAAGAAATAGACATTAATAAAAACTCACTCAGAG 7860
Db 2237 -----AspIleGluLysArgSerAsnSerThr 2245
QY 7861 AAGCGTGCCCAATGAATCCAAATATGTGAT-----ACTATGAATATATACTTCCGCTGAT 7914
Db 2246 GluAspValThrMetSer---ValSerAspSerAsnThrAsnGluLysPheTyrAsp 2264
QY 7915 CTGGGTGACATAGTATTAGAGACAGATATGTTAGAAATGCTGGTACTTACCTCCGCTA 7974
Db 2265 LeuAsnAspCysIleLysAla----- 2271
QY 7975 GAAATTAATTAATATAGGTTTGTGATACATATATGAAATGAGAAATAAATAA 8034
Db 2272 -----GlyIlePheLysGlyLe----- 2277
QY 8035 GGTAGAAATTAATACACGATGTACAAACGTTTCGTTCTGCTGGATGCTAATAGA 8094
Db 2278 -----Arg 2278
QY 8095 AAGATATTTGGAAGCAATGACGTGCAGAACACAGAGATGCAAAACTTTTAGAAA 8154
Db 2279 GluAspValThrPheLysCys----- 2284
QY 8155 GGAAGATGATGATTTGAACGCATACATTA-----ATACAGATTAATGT 8202
Db 2285 GlyIuTyrCysGlyValAspIleCysThrLeuGluLysThrAsnAsnGluLysVal 2304
QY 8203 GGACATAGAGACGATCCACTGCTTGATGATTAATATACCTCAACGGTT-----CGA 8233
Db 2305 SerAlaLysGluAsnAspAsnLysAsnGluIleLeuLeuIleArgValLeuPheLysArg 2324
QY 8254 TGCATGACTGAATGCTGATATATTTATTAAGCACTGATGAGAAATGGAATAATTT 8313
Db 2325 Tripleu-----GluSerPheLeuGluAspTyrAsnLysIle 2336
QY 8314 AAAAAATCATGTGATCTACTTAATAACATGTACAGATGACAGAAATGATTAATGAATAAT 8373
Db 2337 AsnAspLysIleSerHisCys-----MetLysAsnAsp----- 2347
QY 8374 AAGTGTCAACAGCTTAATAAGACATCTCAAGATATATAAATTTTCTCTTAATGAGAA 8433
Db 2348 LysLysSerProCysIleAsnGlyCysGluAsnLysCysAsnGlyValGluLysTrpIle 2367
QY 8434 TCTCTATTCATATACATTAATTAATTAACAAAGAAATGATGAAACCAAAATATATACA 8493
Db 2368 GluLysLysLysSerGluTyrPglLysValArgLysLysGlyr----- 2381
QY 8494 AAAATCTCATCTATGATCATCTTCAAAATTTGTGACAAAAGTTGAAAACCTTTTAAAGT 8553
Db 2382 ---IleAsnGluLys----- 2385
QY 8554 GAATGTTCTGTGAGACCTTTCTGATATCTTCATGAACAAAGTAACTGTTGAATTAAT 8613
Db 2385 ----- 2385
QY 8614 AATTTATGAATAATGATGTTCTCCATATATACAGACATATGCTTTCAGAAACACCA 8673
Db 2386 ---ArgAspLysAsnSerAsn-----GluAlaPheGlu-----Val 2396
QY 8674 AAAAGTATTAAGACAGCTTGACAGTTGATGATCTTCTTAAGAAATCCATTGGATTAATGT 8733
Db 2397 LysSerPheLeuGluTyrIleLeuIleProGluIlePro-----Val 2409
QY 8734 CCTACGATCAATAAACAA-----GATGATGTAAGAAATTAACAACTTTTACCTTCTGCTG 8790

Db 2410 ValThrAspLysGlyLysHisAspSerLeuThrGluLeuLysLysLeuLeuLysCysAsn 2429
QY 8791 AAGAAATGATTAATATATATATCTTGAATTAATGAAACGATACCTGTTCTTAATAGTCA 8850
Db 2430 CysSerGluLysSerGluAsn----- 2436
QY 8851 GATGATTAACAAGAGTGTATTAATGATCTCCAGAGAGACATTAATGTAAGACCTATC 8910
Db 2437 SerAsnGluLysAspVal-----LeuCys----- 2445
QY 8911 ACTGCATATTAATTAAGAAAGGATTAAGAAATTTTAAAAAAAACTTCTTACTTCT 8970
Db 2446 -----LeuLeuLysLys----- 2449
QY 8971 GCTTCAGTCAGACAAATTTAGTCAAAAAATATTAATGGAAGAGAGTGTGCTT 9030
Db 2449 ----- 2449
QY 9031 GAGGCAATGAATATAGTATATGAGATTAATTCGATTAATTAAGAGAACTGATATGATG 9090
Db 2449 ----- 2449
QY 9091 GACACTTCATTAATCTGAAAAAATTAATAATTTTGAACATCAATGAACACCGAA 9150
Db 2450 ---LeuGluAspLysAlaLysAsn----- 2456
QY 9151 AATCGTAAACATGATGGGAAAAATTAATAGACGTGATATGCGACGCTATGTTATGGA 9210
Db 2456 ----- 2456
QY 9211 TATTAATTTGCTACTCAAAAGTAACATTAAGACAGATGCTCAATTAACAAAGAT 9270
Db 2457 -----Cys-----LysAsp 2459
QY 9271 GAAGAAATCAATGATGTTCTTCTGTTGTTAATTAATGCGCAAGCAAGCATGAAGAA 9330
Db 2460 GluAlaSerGlyGlu----- 2464
QY 9331 AAGAAATGATTAATGATTAATTAATAAATAATGCTCTGTCATCAACGAAGATTAATTT 9390
Db 2465 -----ProCysProGlnThrThrSerGluAsnPro 2474
QY 9391 GAAAGCTGAGAAATTAATTAAGACAACTGATGTGCAAGATTAATTAATTAATTAATGAC 9450
Db 2475 AspAspLysAspIleLeuLeuGlu----- 2483
QY 9451 TTGAATATTAATTAATAAATAATTAATAAATAATTAATTAATTAATTAATTAATTAAT 9510
Db 2484 -----GluAsnProValGluAlaProAsnIle----- 2492
QY 9511 GATCAATCTTCAGGTAATTAATTAAGACAAATTAACATCTGAAGAAATGTTCAATATATA 9570
Db 2493 -----CysProLysValGluGluProGluProVal 2503
QY 9571 AAATCAAAAGATTTCTCAATGCGCTTGGAGTTAAATGATTAATTAATTAATTAATTAAT 9630
Db 2504 ---GluGluGluLysCys-----AspLeuAlaGluAlaProSerLys 2516
QY 9631 ACAAAATATATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9690
Db 2517 GluSerSerThrGluGluAsnSerGlyGlu----- 2526
QY 9691 TATTTGTGAAGATGAACACACAAACAAATCATGTAATGGAATTAATAAGAGAA 9750
Db 2527 -----GlySerAsnSerGluGluAsnProArgSerLysProGlu 2539
QY 9751 GAGCAAAACAGTGTCTCAAGACACTATTTCTTACACCCCACTGATGATCTTCTAT 9810
Db 2540 GluGluProProProProThrSerSerGluThrAspThrPro----- 2553
QY 9811 CAAGCAGCTTATCTCAACACATGAGTACCAATATGAT---CCTAAATAAT----- 9861

Db 2554 ---ProfolalaproProthrillegInProserGlnlaaspGlnProthraSerile 2572
 Qy 9862 ---GATATATTGAAGAAAGTACTATCTGTTATTGTATCGGAGTAGGTGATAGCG 9918
 Db 2573 SeraspIleuSerSerThrIleProPhcGlyIleAlaIleAlaLeuThrSerIleVal 2592
 Qy 9919 CTTCATTTCATGAGAAAGAAATTCAAATCGCTGTGGACTTGTGGTATATGAAATATC 9978
 Db 2593 PhelPheLeuLeuLysLysLysThrLysSerSerValaspLeuLeuArgValLeuAsnIle 2612
 Qy 9979 CCCCAGAGAGATGAGTAAGTCCCTAGCTTGGAATCCAAAATGAGTACATACATATAGA 10038
 Db 2613 ProLysGlyIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIu 2632
 Qy 10039 AGTGGTCATATTAAGCAAAACATATATATATATGGAAGAGATAGTATGAGATGAA 10098
 Db 2633 SeraspArgLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLys 2650
 Qy 10099 GATAAATATATGTGGAGCTTATCTCTGATATATCTTATCTATCGAAGAGATGAA 10158
 Db 2651 GlyHisIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLys 2668
 Qy 10159 GAATGGATATATATGAT 10218
 Db 2669 GluLeuaspIleAsnaspIleArgLysIuLysIuLysIuLysIuLysIuLysIuLys 2688
 Qy 10219 GATAGTACTAGAACATCAAAAGGATATACCAAGTAT---GATACACCAAGT--- 10269
 Db 2689 ValValIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLys 2708
 Qy 10270 ---AATGATACACAGCTAGCAATAGATATATATATATATATATATATATATATAT 10311
 Db 2709 ThrGlnAsnaspIleGlnAsnaspIleProSerSerLysIleThrAspAsnGluTyr 2728
 Qy 10312 AATGACATGAACATGATTTTGTATCTCAATATATATATATATATATATATATATATAT 10365
 Db 2729 AsnThrLeuLysaspIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIu 2748
 Qy 10366 ---AATATATCAAAAGCTCAATATATATATATATATATATATATATATATATATCT 10422
 Db 2749 ProAsnaspIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIu 2768
 Qy 10423 GATTAATCTCGAAGAAACCTTTATATATATATATATATATATATATATATATATATAT 10482
 Db 2769 HisAsnValGluGluLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLys 2788
 Qy 10483 AAGAAAT 10542
 Db 2789 GluGluIleAsnIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLysIuLys 2807
 Qy 10543 GCTAGAAATGATTTCTATAGAGCTATATATATATATATATATATATATATATATAT 10601
 Db 2808 ArgaspAsnAsnValLysSerClyIleAspIleuLysIleAsnaspIleValLeuAsn 2825
 Qy 10602 CCTATGAT 2828
 Db 2826 AspTyraspIleTyraspGluValLeu 2834
 RESULT 3
 T28431
 Variant surface protein 1 homolog A4VAR - malaria parasite (Plasmodium falciparum) (frag
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence-revision 15-Oct-1999 #text-change 09-Jun-2000
 C/Accession: T28431
 R:Smith, J.D.; Kyes, S.; Craig, A.G.; Fagan, T.; Hudson-Taylor, D.; Miller, L.H.; Barruch
 Mol. Biochem. Parasitol. 97, 133-148, 1998
 A:Title: Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte
 A:Reference number: Z20486; M01D:99094502; PMID:9879893
 A/Accession: T28431
 A:Status: preliminary; translated from GB/EMBL/DBD
 A/Molecule type: DNA
 A:Residues: 1-3026 <SMT>
 A:Cross-references: EMBL:L42244; NID:93540144; PID:93540145; PIDN:AAD03351.1

C:Genetics:
 A:Gene: var
 A:Introns: 2906/3
 Alignment Scores:
 Pred. No.: 8,256-190
 Score: 4010.00
 Percent Similarity: 42.808
 Best Local Similarity: 29.768
 Query Match: 20.61%
 DB: 2
 Gaps: 127
 US-10-087-013-1 (1-10628) x T28431 (1-3026)
 Qy 88 GAAAGTCACAAAGTCAGAAATGTTTGGAACTGATGCCAAAATATA---AGACAT 144
 Db 13 GluaspAspLysaspAlaLysHisValLeuaspSerIleGlyLysValLysLysGlu 32
 Qy 145 CCATCAAAATATATGCAAAAGAACATGTGATCTGTGAAAGGAGATTGACGAAAGCAAA 204
 Db 33 ValLysLysValAlaAlaAsnLysSerGlnLeuLysGlyThrLeuSerAsnAlaIle 52
 Qy 205 TTTCGTGGTGTCTCTTACCGCACTAAATAGCAATATATATATATATATATATATATATAT 264
 Db 53 PheGluaspGluProLysGlyGlnGlnThrGluAsnAsp-----ProCys 67
 Qy 265 AATTAGATCATTAAGAACAT 324
 Db 68 LysLeuLeuTyrGlyThrGlyHisThrAsnValThrLysGlyHis---GlyArgGluHisPro 86
 Qy 335 TGCATGTGTAGAAACAAACCGATTGATGATGATGATGATGATGATGATGATGATGATGAT 381
 Db 87 CysArgLysGlyThrGluLysArgPheSeraspAlaGlyGlyGlyGlyGlyGlyGlyGlyGly 106
 Qy 382 AAATATGCTATTTTAAAGAAATATGATCTTATAGCTGTGCGCACTAGAAAGCA 441
 Db 107 LysIleLysasp-----SerLysAsnaspGlyLysAlaIleProTyrArgArgLeu 124
 Qy 442 CATATGTGTGATTAATAAAGCTGGAAGCTCTAAATGATATATATATATATATATATATATAT 501
 Db 125 HisLeuGlyValArgAsnLeuGlnAsnIleSerAlaLeuaspLysIleAsnaspThr 144
 Qy 502 TTATTTGGAAATGTACTAGTATACCAAAATACAGAGTGAATCATGTTTAAATATCAT 561
 Db 145 LeuLeuAlaaspValCysLeuAlaAlaLeuHisGlnGlnGlnGlnGlnGlnGlnGlnGln 164
 Qy 562 CCACATATAA-----GGAACCTCAGACGCTTACTGCTCTT 597
 Db 165 ProLysTyrGlnAlaGlnTyrAlaSerSerPheSerProSerGlnIleCysThrMetLeu 184
 Qy 598 GCACGAAGTTTGCAGATATAGTATATGTAAGAGATAGATATGTTTAAACCAAT 657
 Db 185 AlaArgSerPheAlaAspIleGlyAspIleIleArgGlyLysaspLeuTyrIleGlyAsn 204
 Qy 658 GTCATGACAAA-----GTAGAAAGCGGTCTCCGAGAGTTTCAAGAAATATCATGAT 711
 Db 205 LysLysGluLysLeuaspLeuaspLeuaspLeuaspLeuaspLeuaspLeuaspLeuaspLeu 224
 Qy 712 GGAATGAGAGAT---GAAGTAAAAAATGATTAACATCCGATGATCGAATATATATAT 768
 Db 225 LysLeuThrAspProArgAlaLysAspHisTyrLysAspGluProAspAsnAsnPhePhe 244
 Qy 769 AAATTAAGACAGACCGTGGAAATGCAATATAGAAATAAAGTATGGAAGCTATATATATAT 828
 Db 245 GlnLeuArgGluaspTrpTrpAsnAlaAsnArgGlnGlnValTrpLysAlaIleThrCys 264
 Qy 829 GATGCATCATATTAATCTGATATTTTATGCAATCGAAAGTAAATACACATATTTTCA 888
 Db 265 HisAlaGlyLysSeraspLysTyrPheArgLysThrAlaCysGlyThrGlyThrGlyThr 284
 Qy 889 AATCCTAATATGCGGCATATAACAGAAAG-----GTTCTACG 927
 Db 285 GlnGlyArgCysArgCysAsnAspAspLysLysProGlySerAsnThrAspProProThr 304

Db 971 AsnLeuValThrIlePheGlyLysIleLysValGlnArgLysGlyIleAspThrSerLys 990
QY 2962 TATAATGATGATGCCCCCAATATTTAAATTTGAGGAAATTTGCGAGGAAATGAGCAATAGA 3021
Db 991 TylThrAsnThrAspGlyLysHisAsnGlnLeuArgGlnAspTrpTrpGlnAlaAsnArg 1010
QY 3022 GCCAAATATGGGAGCCAGCAATGAAATGTATATAAATATTTGAAGATTAATCGGCACAC 3081
Db 1011 ArgGlnValTrpLysAlaMetLysCysAlaLeuLys 1022
QY 3082 CAATCAACACAAAGTACTTATTTGGGATATAGATATACCATTCATGATATATATC 3141
Db 1023 -----GlyGlnLysIleAsnCysGly-----AlaThrProLysAspArgTyrIle 1037
QY 3142 CCACAAAATTAAGATGATGACCGCAATGGCAGAAATGTACTACGAGTGCAGAAATAGC 3201
Db 1038 ProGlnArgLeuArgTrpMetThrGlnTrpAlaGlnTrpPheCysLysGlnGlnSerArg 1057
QY 3202 GAGTATGATTAAGTTGAAGGAAAGTGTAGAGCTGTAGAGATTAAGATTAAGTCAAGGC 3261
Db 1058 LeuTyrAspLysPheLysMetGlyLysCysGlyIleCysMetAsn-----Gly 1072
QY 3262 TGTACAAAGAGAGGTGATCAGGTGTACAGAGTGCAGAGCTGTATGATTAATATAT 3321
Db 1073 IleCysAsnLysValLysAspAspCysAlaLysCysThrGlnAlaCysLysGlnTyrLys 1092
QY 3322 GATATATAGATTAATGAAAGCAATGAAATATATATATAGATTAATATATC-----AAA 3375
Db 1093 ThrLysIleGlnProTrpLysAspGlnTrpGlnLysLeuGlnLeuGlnLysAlaLeuSer 1112
QY 3376 GAATTCATGAAACACACAAATGTCTGTATATAGTATGATTAAGCTTCCACTACT 3435
Db 1113 TyrLeuHis----- 1115
QY 3436 GCCAAATAATCATATACACAG-----AATGTT 3462
Db 1116 AlaLysAsnAspSerArgTrpMetAlaPheGlyLysThrAspProAspTrpGlnGlnVal 1135
QY 3463 ATTGAATTTTGTCCGATATATACCAACA-----AATGTCGCAAAAGTAAATGAAGTGT 3519
Db 1136 ValHisPhePheLysGlnLeuGlnGlnAlaIleLysSerSerThrSerLysArgProLys 1155
QY 3520 ACTAGTATGAAGTGTCTCATGTGTACTACACAGAGTGAATTAAGTGTGACAGATAT 3579
Db 1156 ArgSerThrAspAlaIleThrThrAspProThrTrpProLysSerThrAlaIleGlyTyr 1175
QY 3580 CTCATGATACAGAAATTTGATGATGTCTAGTCAACAATGAGTTTGTGATGAAAAA 3639
Db 1176 IleHisGlnGlnIleGlyAsnAlaGlyCysGlnIleGlnLysHisPheCysAspAsn 1195
QY 3640 AGTATGATGAAGATTAACGAAATAATGCTTAAAGATTAACCCACAGACCATGATGTT 3699
Db 1196 LysAsp-----LysTyrValPheArgGlnLysProLysAspHisAspGln 1210
QY 3700 GCGTGTGTTGTAAGTGAATGATCGAAACGACAGAGGTACAGATTAACGAAAAA 3759
Db 1211 AlaCysAsnCysThrGlnAsnValGln-----LysProLys 1222
QY 3760 GCGGAGAAAGATACGGA-----TGTAAACAGTGAATGATATCTTAAAGAA 3810
Db 1223 LysGlnGlnIleAspThrLysLeuAspValCysAlaIleValLysThrAlaLeuThrThr 1242
QY 3811 AACGATGAAGAAAGCAAGTGAAGATGTGTATCCAAATAAAGATAGTAATGATATCCC 3870
Db 1243 GlnAspAsnLeuThrLysAla-----CysGlnGlnLys-----TyrGlyHisPro 1257
QY 3871 GAT-----TGGCAATGC-----GGAATATATAATTTAGTGAAGACCCCT 3909
Db 1258 GlnArgHisTrpGlyTrpLysCysIleSerGlnThrThrThrLysSerSerAspSerGly 1277
QY 3910 CGTGTGTATGCCCCCAGAGACAAAGTATGCGTACATTTCTTGGCAAT----- 3963
Db 1278 SerIleCysValProProArgArgLysLeuTyrValThrProLeuThrLysTrpAla 1297
QY 3963 ----- 3963
Db 1298 GlnGlnAlaThrGlnGlnLysProThrSerProGlnAlaGlyGlyGlnAlaThrLeuProThr 1317
QY 3964 -----GATATGAAATTAATAAATAATTAACATCAATCAAGTTAATTAATAAGCT 4011
Db 1318 ProAlaThrAlaSerSerGlnAlaProLysGlyAspSerLeu-----LeuLeuThrAla 1335
QY 4012 TTCATCAAAATTCGACACAGCAAAACATTTCTGTATGATTAATTAATAAGTAAAGAT 4071
Db 1336 PheIleGlnSerAlaIleAlaValGlnThrPhePheLeuThrHisLysTyrLys-----MetAsp 1354
QY 4072 GGTCAAGCAAAATGAACTGCATTAAGATTAATAAGAGCAAAATTCCTCCGCAATTTTG 4131
Db 1355 AsnAsnGlyLysAspAlaLysAspLysLeuLysLysGlyGlnIleProGlnGlnLys 1374
QY 4132 AGATCCATGTTCTACACATTTGTGAGATTAATAGATTTTATTTTGAACA----- 4182
Db 1375 ArgGlnMetPheTyrThrLeuAlaAspTrpArgAspIleCysIleGlyValLysGlnAsp 1394
QY 4183 -----GATATATCAAAAGTCAATGCTGAGGAGCAATTAATAAGAG----- 4224
Db 1395 ValIleLysAlaLeuGlnAlaLeuAlaSerSerAspAsnLysSerGlyAsnAsnIleLysGlnLys 1414
QY 4225 -----CAATGATTCCTCTTCAAAAATGCTGACCAAAAATCTCCATATGA----- 4272
Db 1415 SerAspLysIleGlnAlaIleLeuLysGlnSerLysSerLysLeuProGlyLysLeuPro 1434
QY 4273 -----AAAAACGCAAGAAATGCTGACAGAAACATATGATGATGATGAT 4314
Db 1435 ValThrProAsnAsnValLysAsnProLysThrTrpTrpAspGlnAsnAlaHisIle 1454
QY 4315 TGGGAGCATGCTATGCTGACCTACTA-----AAATTTGGGCAAAA---AAA 4359
Db 1455 TrpHisGlnMetValCysAlaLeuThrTyrLysGlnAspThrGlnGlyAlaLysGlyLys 1474
QY 4360 GATGATTTTACCGAAAC----- 4377
Db 1475 ThrSerIleThrGlnAspProThrAlaIleGlyLysLeuThrAspAsnAspGlyLysLys 1494
QY 4378 -----TACGTTACACAAAGTCAATTTACT----- 4404
Db 1495 ProLysGlnAspLysTyrAspTrpLysIleValLysIleSerSerValProSerSerAsn 1514
QY 4405 -----GACAAACGACACTTTGGAGCAATTTGCCAAACGACCCCACTTTTACGATG 4458
Db 1515 ValProSerGlyAspThrLysLeuGlnGlnPheSerArgArgProThrPhePheArgTrp 1534
QY 4459 CTAACGAAATGATGACGAGCACTATGCTATACAGCAAAATAATTTGAAGATGTGCAG 4518
Db 1535 LeuGlnGlnTrpGlnGlnGlnPheCysArgGlyArgLysAspLysLeuGlnIleGln 1554
QY 4519 GAAATAATGTAAG-----TCMAATGACCAATTTG 4545
Db 1555 LysGlnLysTyrGlyLeuAsnTyrArgGlyHisArgGlyLeuTyrCysSerGlyAspGlnHis 1574
QY 4546 AAGTGTGAT-----ACGAA 4560
Db 1575 IleCysGlnLysThrAspThrSerArgAsnAsnThrPheIleAspLeuHisCysProArg 1594
QY 4561 TGTATATAGAAATGCGAGCACTAGTAAATATATG-----AAAAAATAAAGATGATG 4617
Db 1595 CysLeuLysGlnCysIleLysTyrLysArgTrpIleGlnLysLysGlnLysGlnPheHis 1614
QY 4618 CCACAAATATAATTTCAAGGATGAACGCGCAAAAAAAGATTCGATGACACACATTT 4677
Db 1615 AsnGlnLysAsnAsnTyrGlnLysGlnPheAsnAsn-----AspLeuLysGlnLys 1631
QY 4678 GGTGTATGTTACAGACTTACTGGAACGATGCAACAGATTAATCTTGAACAGAAATTT 4737
Db 1632 Gly-----TyrSerSerPheAsn-----AspPhe 1639

QY 4738 ACTGCTAGTGTGTGAT---AACCTGGAATGCTCTGTGTGACAAAGAT---ATA 4791
|||||
Db 1640 LeuAlaSerLeuAsnHisCysLysHisGluHisAsnArgAspLysAsnLysIle 1659
|||
QY 4792 CAATGTGTACAAAACAGGCTTACTATGATGCGGACAAACATGT---GGCTGCACAAAA 4848
|||
Db 1660 GluPheAsnAsnHisThrLysThrPheGlyProSerGluTrpCysLysAlaCysProVal 1679
|||
QY 4849 TTTATTTGAAAATGACGACAAATATTAACATTTTCAGTAAGATTAAGTGCACAAAGATTA 4908
|||
Db 1680 TyrGlyValLysCysAsnLys-----LysAsnGlyGluCysGluThrIle 1694
|||
QY 4909 GTTAAGAGGACAAACACAGCTGCTATTAGTGCACAAAACAGGCTCTTAATTAATCAAT 4968
|||
Db 1695 HisLysThrAspLeuAsnGly-----GluAsnAspAsnAsnTyrThr 1708
|||
QY 4969 AACCTGAAAAGATTGACTGAAGATGTGCTTTTCTCTCTGCTGACACTATATGTTT 5028
|||
Db 1709 AspIleLysValLeuValIleAsp----- 1716
|||
QY 5029 CATGCAATTGGATGCAATTTATACAGATCCAGATTAAGATGAAAATGGCTTCCGAAAA 5088
|||
Db 1717 -----ArgLysGlyGluSerAsnAspGluGluLeuLysAsnValCysAsnAsnThrSer 1734
|||
QY 5089 AGATTGATGAAGTGGCGGCGACGGAAGGTACAAATTTGGGTCAATATCTCAAGAAAA 5148
|||
Db 1735 LeuPheLysAspSerSerVal-----GlnTyrTrpLysCysGln 1747
|||
QY 5149 AAAGAAAAAGAG-----AAATAAAAAAGTCGGATGGGCACAAA 5187
|||
Db 1748 LysLysAsnGluValAspGlnCysIleIleAspAsnPheLeuAspIleAspLys 1767
|||
QY 5188 TATTTCTTATGAGTCCCGCTGTAGTGTGTAATATATGTTTATGATTAAGAGAT 5247
|||
Db 1768 Tyr-----MetGluPheAsnValPhePheGlnArgTrp 1778
|||
QY 5248 ATATTTCTAGCTATTGTAATTTTGGAAAGATGAAAACAAAACCGGAAAAATTGGAAG 5307
|||
Db 1779 LeuArgThrPheValHisAspTyrAsnIleLeuLysAspLysIleLysProCysIleLys 1798
|||
QY 5308 AAATATTTTAACAAAATGACATCAGTGGCAAGGAGTGTACTATACAGAAAT 5367
|||
Db 1799 -----ThrLysAspGluLysSerAsnLysCysIleAsnGlyCys 1811
|||
QY 5368 CCGGATAGTACTCGCGCAAAATTTTCTGGAACGAAAATTAAGAAATCTGTGTGAACGA 5427
|||
Db 1812 LysGlyLysLeuGluCysValLysLysTrpLeuLysGlnLysGlnAspGluTrpLysAsn 1831
|||
QY 5428 ATGATATGCGGGTACAAACCTGCTAGGAGTATGGAATAGTGA----- 5472
|||
Db 1832 IleLysAspHisThrGlyLysAsnLysSerLeuTyrGlyIleProHisThrVal 1851
|||
QY 5473 -----AATAGTCAAGAGTATGAGAGATCTAATAAAATGCTGTCTGTA 5517
|||
Db 1852 LysSerThrPheValGluGlnLeuTyrPheAspLysAspTyrLysLysAlaGlnAspVal 1871
|||
QY 5518 CCTTCAGATGATGATTAATCTATGGGAAAAAATCGGATGAAGTACTGCTATCAATTT 5577
|||
Db 1872 IleGluAsp----- 1874
|||
QY 5578 CTTCGATGCTTTCGCCAATGGGTGAAGATTTTTCGAAACATAAAGAAAAGAAATTGGAG 5637
|||
Db 1875 -----GluAsnGluArgLys 1879
|||
QY 5638 AAATTTGATGGGCGTATATGATTAATCTTGTGCTATATGAAAGTAAAGAAAGAAA 5697
|||
Db 1880 LysIleTrpGlyCysThrAspGlyValGluCysThrAsnGluGlnTrpLysGluAsnLys 1899
|||
QY 5698 TGTACAGATCGTGTACACAAATTAATAAAATTTATAGTACGAAACACACATATGAA 5757
|||
Db 1900 -----AspPheIleThrAsnLeuIleLysLysLeuGlnIleLysIleGluSerCysGln 1917
|||

QY 5758 AAACAAATCAAAAATATGTGTGAGATTAAGACAAAATATATTTCCAGACATCTGTGGCA 5817
|||
Db 1918 ThrGlnHisAsnProAsnGly-----LysThrLysProCysAspGluIleProPheHis 1935
|||
QY 5818 AAGATGACACAGACCGGTCCGATATTTAGACAAACAAATTAATAAAATTTGTGAAT 5877
|||
Db 1936 SerAsp-----GluThrLeuAspGluIleThrAspThrThrAspAsp 1950
|||
QY 5878 AAAGTGCAGATTTGTGAATATAG-----TGATGAAGAATGTGTCCACA 5922
|||
Db 1951 AspAsnSerAspLysIleTyrAspThrLysProPhePheCysProLysAspValGluAsp 1970
|||
QY 5923 CACCGATTAACGTATGTGTAATAGTCAAAATATGCCCCGATCATTAAGCATGAACAAA 5982
|||
Db 1971 ThrLysGluThr-----GluLysProLys 1978
|||
QY 5983 GAAGTTGAAGAAAGTGTATATGTAAGTCCAGACGAGTCCACACAGTGTACGAAGGAA 6042
|||
Db 1979 -----ValLeuProGlyProPro----- 1984
|||
QY 6043 ACACCGTACACACGCGGTATCTGATATCAAAAACGACGCAATGAAAAAGACGAAA 6102
|||
Db 1984 ----- 1984
|||
QY 6103 ACAGCCCGCTTACAAAACACCGGAAAAAGTGAATACTAACACGAATGCGAGCA 6162
|||
Db 1984 ----- 1984
|||
QY 6163 CAACAACGAAACCGACAGACGACACAAACAAACGAAACGAAATCAACAGCAACACA 6222
|||
Db 1984 ----- 1984
|||
QY 6223 ACAGATCTGACGTGGGACAAATGTAAAGCCATTTCTGATTAACACGATAGCAGG 6282
|||
Db 1985 -----AspLacCysGluIleValGlyGluIleLeu-----AsnGlyAsnAspGlyThr 2000
|||
QY 6283 GGTGGAATAGAGGCTGTATCCAAAACGAT-----GCACAAATCTCTAATGG 6333
|||
Db 2001 LysLysIleGluLysCysAsnThrLysTyrTrpThrLysAsnAspTyrProGlyTrp 2020
|||
QY 6334 GGTGTATGTAGTAAAGTCAAAAGAAATGAAATGCGCATGTATGCTCCCTAGAGCA 6393
|||
Db 2021 AsnCys-----ThrAspLysValIleAsnArgGlnGluLysCysMetProProAlaGly 2039
|||
QY 6394 AAAAATATGATATTAATAATATATATATTAATTAATTAATTAATTAATTAATTAATTA 6453
|||
Db 2040 GlnLysLeuCysIleHisAsnLeuGlnHisLeu-----SerGluLysAlaThrGlu 2056
|||
QY 6454 AATGATATTAAGAAGCTTTATTAATTAATGTCAGCAATAGAACTCAATTTTGTGGTTA 6513
|||
Db 2057 ThrGluLeuArgLysAlaPheIleGluCysAlaIleGluThrPheTrpLeuTrpAsp 2076
|||
QY 6514 AAATAT-----ATAATGAATAAT 6531
|||
Db 2077 LysTyrLysGluAspLysLysAspGluLysLysThrGluLysGlyIleSerAspAsp 2096
|||
QY 6532 CCT---GCACGAGAAAATGAAATTCGAAAATGAAACAAATTCAGATGAAATTAAGAAATA 6588
|||
Db 2097 ProAspAspProGlnLysLysLeuGluGlyThrIleProGlnAspPheLysArgGln 2116
|||
QY 6589 ATGATATATCATATGTGTATTAAGATATGTTTGTGAACGTATATTTCTATGAT 6648
|||
Db 2117 MetPheTyrThrTyrGlyAspTyrArgAspPheLeuPheGlyThrAspIleSerLysGly 2136
|||
QY 6649 AAA---AAATTTAATCTGTAACAAATAGTGAACAAACCAATTCATGAATAATTAAG 6705
|||
Db 2137 HisGlyLysGluSerAlaLeuGlyLysLysIleAspSerLeuPheLysAsnGlyAspGln 2156
|||
QY 6706 AAAAACACGATTAATAAAAGATGAAGATTAAGTAAATATTTTGGCGCAAAATTAAG 6765
|||
Db 2157 LysSerProSerGlyLysThrProThrGlu-----TrpTrpAsnAspTyrGly 2172
|||
QY 6766 AAATTTATTTGGAGAGATGATATATGATTAATCTATCATCTCACAGACGAAACGAA 6825
|||


```

QY 8923 TATAGAAAGGTGATTAAGAAATTTTAAAAAACTTCTACTTCTGCTTCAAGTCAA 8982
Db 2809 -----Ser 2809
QY 8983 GGACAATTGTTAGTCAAAAATATTAATGGAAGAGATTGCTTTGAGCAATGAA 9042
Db 2810 ProserleuenuilypProserPhe ----- 2817
QY 9043 TATAGTTATGACAGATTATTCGATATTAATTAAGAACTGATATGACACTTCATTA 9102
Db 2817 ----- 2817
QY 9103 TCTGAAAAAATTTAAAAAATATTGTAACATCAATGAGCAACGAAAAATCGTAAACA 9162
Db 2817 ----- 2817
QY 9163 TGGTGGAAAAATTAATAGACTCAGATATGCGACGCTATGTTATGTGGATTAATAATTGCT 9222
Db 2817 ----- 2817
QY 9223 ACTTCAAAAGTACATTAGATGAGAGATGCTGTCATTAATCAAGAGATGAGAACTAAT 9282
Db 2818 -----Pheleuprollys ----- 2821
QY 9283 CAGTTCTTCTGTTGCTTAATGATGGGCAAGCAACATGTAAAGAAAAACAATGTA 9342
Db 2821 ----- 2821
QY 9343 AGTGATTCAATTAACAAAATGTCCTGTTCAAGCAAGATTAATTTGAAAGGTCAAGAA 9402
Db 2822 -----LysProllys ----- 2824
QY 9403 TTATTAAGACACCTGATGTCAGATGATTAAGAAATATATAGCTTGAATATATG 9462
Db 2825 -----MetlyserhetyrProllysileglyilegly 2834
QY 9463 ATAAAAATATCAATGGAATCTTAATATTAATAATATAGCAATTAAGATCAATCTTCA 9522
Db 2835 ValLeuHisProPheileAsnMet ----- 2842
QY 9523 GGTAAATATAGCAATTAACCATCTGAAGAAATGTTCACTCATATATTAATCAAAAAGAT 9582
Db 2842 ----- 2842
QY 9583 TCTCAATGCGCTTTGAGATTAAATGATTAATGAATAGTACAGAACAAAAATATAT 9642
Db 2843 -----ValAlaAspProilethrlleHisgluThrValAlaLysThrPheAsnAsn 2859
QY 9643 GAAAAATATGATTCAAAGAAAGTACTRAAAAAATATATATCTGCTTATATTTGTTGAA 9702
Db 2860 AlavalProglInphe ----- 2864
QY 9703 GATGAAGAACACAAAATATCATGTACTAGATGAAATTAATAAGAGAGCAACAGT 9762
Db 2865 -----HistIleAsnProAspLysThrAspVal 2873
QY 9763 CGTCTTAAGCACTGATTTCTTTACACCCCATGTAGATTTCTTCAATCAAGCAACCTTTA 9822
Db 2874 AlAProPro -----ThrLysAsnIle ----- 2880
QY 9823 TTCTCAACACATGAGTAGACAAATATGATCTTAATAATATATTTGAAAGTAGATATC 9882
Db 2881 -----LeuAsnIleValLeuProSerAlaIle 2889
QY 9883 TCTGTTGTTATGATCGCGCTTAGTTCATGATGATGATTCATTTCAAGAAAAATTC 9942
Db 2890 ProValAlglyIleAlaLeuAlaLeuGlySerIleAlaPheLeuPheLeuLysLysLysThr 2909
QY 9943 AAATCGTCTGTCGATTTGCTTATATGCAATATCCCGCAAGAGAGATGATGAAATGCTT 10002
Db 2910 LysHisProValAspLeuPheSerValIleAsnIleProLysSerAspLysIlePro 2929

```

```

QY 10003 ACCTTGGAATCCAAAATATAGTACATACATATAGAGATGCTCCATTAAGCAAAA---- 10059
Db 2930 ThrLysLeuSerProAsnArgLysIleProTyrThrSerLysLysIleValGlyAsnGly 2949
QY 10060 ACATATATATATATGGA-----GGAGATACATAGTGCAGATCAAGCAATATATATG 10113
Db 2950 ThrPheThrLeuLysgluIleValIglyThrAspSerGlyThrAspHisIle----- 2967
QY 10114 GACTTATCTTCTCTGATATATTAAT-----TCATCCGAAAGTACATGATGAGAAATG 10170
Db 2968 -----SerAspIleThrSerSerGluSerGluSerIleThrIleValIleValIle 2983
QY 10171 AATGATATATATATGATACAGTACCTTAATATTAATAATGATGATGATGATGATG 10230
Db 2984 AsnAspIleTyrHisIleValLeuLeuAsnIleLysThrLeuIleGluValIleGlu 3003
QY 10231 CCATCAAAAAGG-----GATATPCCAAGAT-----GATACCAACAGTAAT 10272
Db 3004 ProSerGlyLysLeuSerGlyAsnThrIleProThrSerGlyLysAsnThrProSerAsp 3023

RESULT 4
T28432
Variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28432
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfelde, J.A.; Peterson,
Cell 82, 89-100, 1995
A>Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: Z20487; MID:95330813; PMID:760578
A:Accession: T28432
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3078 <SUX>
A:Cross-references: EMBL:LA0608; NID:g886374; PID:g886375; PIDN:AAA75396.1
C:Genetics:
A:Gene: var-1
A:Introns: 2611/3

Alignment Scores:
Pred. No.: 8,48e-165 Length: 3078
Score: 3501.00 Matches: 1057
Percent Similarity: 38.65% Conservative: 470
Best Local Similarity: 26.75% Mismatches: 934
Query Match: 18.00% Indels: 1490
DB: 2 Gaps: 147

US-10-087-013-1 (1-10628) x T28432 (1-3078)
QY 25 ATTATTAATAATGGGATGCGATCATCATCTTATGAGAGAGATGCTTAAGCCCTATTATA 84
Db 4 LeuAlaLysMetGlyProllysGluAlaIleGlyLysAspIleGlyAsp ----- 20
QY 85 AAAGAAAGTCAAAAAGTGCAGAAATGTTTGAACGTTATGCCAAAAATATTAAGCAT 144
Db 21 -----GluSerAlaLysHisMetPheAspArgIleGlyLysAspValIleTyrAsp 36
QY 145 CCATCAAAA-----TATGCAAAAGAACATGTGATTCGTTGAAGGAGATTTGAGCAAGCA 201
Db 37 LysValLysGluLysAlaLysGluArgLysGlyLeuGlnIleValGlyLeuSerGluAla 56
QY 202 GAATTCGTCGTCGCTCTTACGCCAGTAATTAAGCATATTAATTAATTCATATCA 261
Db 57 LysPheGluLysAsnGluSerAspProGlnThrProGluLys -----Pro 71
QY 262 TGTAAATTTAGTATAGTAAGCAATCAATTAATTAAGGATGATGATGATGATGATGATG 321
Db 72 CysAspLeuAspHisLysIleHisThrAsnVal-----ThrThrAsnValIleAsn 88
QY 322 CTTGCGCATGTCGTGAGCAAAAACCGATTTGATGAAGATGAGAAATCTGAAATGCT---GGA 378
Db 89 ProCysAlaAspArgSerAspValAlaArgPheSerAspIleLysIleGlyGlnCysThrHis 108

```

```

QY 379 AATAAATACGTAATTTTAAAGAAAAAGANGCTATAGCCGTGCGCAGCCTAGAGA 438
Db 109 AsnArgIleLysAspSerGlnGlnIlyAspAsnLysGlyAlaCysAlaProTyrArgArg 128
QY 439 CGACATATGCTGTAATAAACTGGAGCCTCAATATATTAATACCCAAATATTCAT 498
Db 129 LeuHisValCysAspGlnAsnLeuGlnIleGluProIleLysIleThrAspThrHis 148
QY 499 GATTATTTGGAAATGACTAGTATACGCAAAATACGAGGTGAATCATTTGTAATAT 558
Db 149 AsnLeuLeuValAspValCysMetAlaIleLysPheGlnIleGlnIleThrGlnAsp 168
QY 559 CATCCA---CATAAGAACT-----TCAGACCTTGTACTGCTTTCGA 600
Db 169 TyrProLysTyrGlnAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 188
QY 601 CGAGTTTTCAGATATAGTGATATTTAGAGAAATAGATATGTTTAAACCAATATGC 660
Db 189 ArgSerPheAlaAspIleGlyAspIleValAlaArgIlyArgAspLeuTyrLeuGlyAsnPro 208
QY 661 CATGAC-----AAAGTAGAAACGGCTCTCGAGAGGTTTTCAGAAATA 705
Db 209 GlnGlnIleLysGlnArgGlnGlnIleGlnIlyAsnAsnLeuLysThrIlePheGlyLysIle 228
QY 706 CATGAT-----GGAAATGGAAGATGAAGTAAATGATTACAAATCCGATGATCT 756
Db 229 TyrGlnIlyLeuAsnGlnIlyAlaGlyAlaArgTyrGlyAsnAsp-----ProGlu----- 244
QY 757 GGAATTTATTTAAATTAAGAGAACGATGATGATGATGATGATGATGATGATGATG 816
Db 245 -----PhePheLysLeuArgGlnAspTrpTrpTrpAlaAsnArgIleThrValIleTrpLys 262
QY 817 GGTATTAACATGAGCAGCATATATAATCTGATATTTATGCAATCAGAAAGTATACA 876
Db 263 AlaIleThrCysAsnAla---TrpGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 280
QY 877 CCATTAATTTTCAAACTCAATGCGCCCATTAACAAAGAAAGTCCCTCCATTTAGAT 936
Db 281 GlyIleArgThrLysGlnIlyTyrCysArgCysAsnAspAspGlnValProThrTyrPheAsp 300
QY 937 TATGCCCCCATATTTACGTTGTTCCAGCATGGGAGAGATTTTCCGAAAAAGA 996
Db 301 TyrValProGlnIlyLeuArgTyrPheGlnIlyTrpAlaGlnAspPheCysArgLysLys 320
QY 997 AATATTAAATGAAAAGCTCAAGACCTCCGTCTGCT-----AATACAAAGA 1044
Db 321 AsnLysLysIleLysAspValIlyAspArgCysArgGlyLysAspLysGlnLysLysAsp 340
QY 1045 CGCTTATATTTAGTCATTAATGACATGATTTGACAGCACTATTTGAAAAAAGCTAT 1104
Db 341 Arg---TyrCysSerArgAsnGlnIlyTyrAspCysGlnLysThrLysArgAlaIleGlyLys 359
QY 1105 TTGCAATTTGGATTAATAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1164
Db 360 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTrp 379
QY 1165 TTAGGGAATCAACAAGAGACTTTTAAAAAAGAAAAAAGAAATGAAAAAATACAA 1224
Db 380 IleAsnAsnGlnLysGlnIlePheAspLysGlnIlyLysLysLysLysLysLysLysLys 399
QY 1225 TCATAT-----TTATCGAAGATTAACAAATTTGTCAAT 1257
Db 400 LysTyrGlnLysAsnGlnIlyAlaSerGlyLysSerArgGlnLysAspArgAlaGlyLysThr 419
QY 1258 AATATTAATAGTGA---TATTATAACAATTTTAAAAAAGCTTAAAGAAAGCAATAT 1314
Db 420 ThrThrAsnTyrAspGlyTyrGlnLysLysPheTyrAspGlnLeuAsnLysSerGlnTyr 439
QY 1315 GCAACTAATGACACTTTTAAATTTAAATTAATGAAAGAAAGTATGT----- 1362
Db 440 ArgThrValAspLysPheLeuGlnLysLeuSerAsnGlnIleCysThrLysValLys 459
QY 1363 -----AAGAGCATTTACAGAGAAAAAGATATTACTTTACTTACAGCTCT----- 1410
Db 460 AspGlnIleGlyLysThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 479
QY 1411 -----GATGCAAAAGGATATTTTATTCGTTGCAATATTTCCCAAGTGTG 1455
Db 480 SerGlyThrAsnValGlnSerGlnIlyThrPheTyrArgSerLysTyrCysLysProCys 499
QY 1456 CCCGATGCGGGGTCAAAATGTAGTATTAATAATCACACAC---AATCAGATATGAT 1512
Db 500 ProTyrCysGlyValLys-----LysValAsnAsnGlnIlySerSerAsnGln 515
QY 1513 CGTGAACGTGTAAATAT-----GAAGACTATTAACCTCCATGGGCT 1554
Db 516 TrpGlnIlyLysAsnAsnGlnIlyLysCysLysSerGlyLysLeuTyrGlnProLysProAsp 535
QY 1555 GTGAAGCTTACTAATATACATGCTCTTATAGCTGTAATGAAACAGATGATTTACACA 1614
Db 536 LysGlnIlyThrThrIleThrIleLeuLysSerGlyLysGlnHisAspAspIleGlnIle 555
QY 1615 AATTAAGAAATTTTGTGAC-----AGCTCA 1641
Db 556 LysLeuAsnLysPheCysAspLeuLysAsnGlnLysAspThrIleAsnSerGlyLysSerGly 575
QY 1642 ACTAATTAACAAAGATTAATAATAT-----CAAAATGGAATGTC 1680
Db 576 ThrGlyLysSerGlyGlyLysAsnSerLysArgGlnIleLeuTyrGlnIleTrpLysCys 595
QY 1681 TATTTAAGATGAAATATTAATATGATGTAACTGGAACAAATACGAAATCAATATAT 1740
Db 596 ---TyrLysGlyLysAspValAlaValAlaLysValGlnHisAspGlnAspAspGlnIlyAspTrp 614
QY 1741 GATTAATCCTAAG-----ATAATA----- 1758
Db 615 GluAsnValLysAsnAlaGlyLysLeuCysIleLeuLysAsnGlnLysLysAsnLysGln 634
QY 1759 -----TCATTTCAATATTTT 1773
Db 635 GlnIlyLysAsnThrSerGlnLysGlnProAspLysIleGlnLysThrPheAsnProPhe 654
QY 1774 TTTGATTAATGAGGTATACATTTTAAAGGATATTTTAAAGTATGATGATGATGATGAT 1833
Db 655 PheTyrTyrTrpValAlaHisMetLeuLysAspSerIleHisTrpLysLysLysGln 674
QY 1834 ACTGTATA---AATTAATACAAACGACATGT---ATTGATGAATGTATACAGAAATGTC 1887
Db 675 ArgCysLeuGlnAsnGlnLysAsnArgIleLysCysGlnLysAsnLysCysAsnAsnAspCys 694
QY 1888 TTATGTTTTCAGACATGCTTAAACAAAGAAAGAAAGATGATGATTAAGAAATCTG 1947
Db 695 GluLysPheLysArgTrpIleThrGlnLysLysAspLysIleValGlnHis 714
QY 1948 TTCACAAAAAAGATTAATACAG----- 1971
Db 715 Phe---LysThrGlnAsnIleLysGlyArgGlyLysSerAspAsnThrAlaGlnLeuIle 733
QY 1972 -----CAATGCTATTAATTAATTAATTAATTTTGAAGTATTTT 2016
Db 734 ProPheAspHisAspTyrValLeuGlnIlyAsnLeuGlnIleGlnIleLysGlyAsp 753
QY 2017 -----TTTAAAGTTATGATTAACCTTGACAAAGATGAGCAAAATGCG 2058
Db 754 SerGlnAspLysSerGlnIlyLysSerGlnLysSerLeuAspAlaGlnIleAla 771
QY 2059 AAGAACTTAAGAAATATTAATAAGAAAAAATGAGTTTCCAAATTTGGAATATAT 2118
Db 772 -----GlnIleuLeuLysHisLeuArgGlnIleIleLysSerGlnLysAsn 787
QY 2119 AGGAC-----TATTAGAAATGCAATGACACTTGTGTA 2154
Db 788 GlnGlnIleAspValGlyGlyLysValThrGlnIlyLysAsnIleMetAspLysLeuLeu 807
QY 2155 GATCACTTAAAGAAACTGCCAGATATGT----- 2184

```

Db 808 AsnTyrGluLysAspLeuAlaAspLeuGluLeuHisGluAspGluGlu 827
OY 2185 ---AAGACAAATATACAAAGACATGGAACATCCATATGCAACAAACCCG 2241
Db 828 LysGluLysGluAspLysGluGluGluGluGluGluGluGluGluGluGlu 847
OY 2242 TGTGTAAACCTGTGAGCGACGAA---CCGACTAAATATTAAGAAATAGCA 2298
Db 848 CysSerGlyLysSerGluAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGlu 867
OY 2299 TACTTTAAAGAGTCATACGAGAACGCAATCTGCTTCAATTAATTGAAGA 2358
Db 868 MetHisHisLysAlaLysThrGluLeuAlaSerArgAlaLysSerAlaLeuArgGly 887
OY 2359 AAGCACACGAGCTATATTAACGTTGAGGAGGAGAAAGCACTTCAAGCAATTTA 2418
Db 888 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlyGlu 907
OY 2419 TGTAGATTAATGATTAACATCTTAATCTGATCTGTTTCAATGAGCAGATGAT 2478
Db 908 CysLysIleLeuAsnLysSerAspSerArgGlyAsnSerGlyGlyProCysThr 927
OY 2479 GCGAAGCGCAGCTAGTATACAAACAGTTTGTCTAGAACTGAATGGAAGT 2538
Db 928 GlyLys---AspLysPheHisGlyLysAlaArgMetArgIleGlyThrGluTyrSerAsn 946
OY 2539 GATCCGAGACACATGCTAAAGATGATGATTTATATGCTTACAGCAGATGAT 2598
Db 947 IleGluGlyLysGluGlnThrSerTyrLysAsnValPheLeuProPheArgGluHis 966
OY 2599 ATATGATCATCAATTTGGAACATTTACAAAGCATGATGATCCACTTAATGATAT 2658
Db 967 MetCysThrSerAsnLeuGluAsnLeuAspVal-----GlySerVal 980
OY 2659 GTT---GATGATTTAGTATTAATCTCTTTTGGGAGATCTTCTCATAGCAAAATAT 2715
Db 981 ThrLysAsnAspLysAlaSerHisSerLeuGluLysPheValGluLeuAlaLysThr 1000
OY 2716 GAAGCAAAACAGATTAATACGATGATTAAGAAAGAAATTAACCTAAAGGCCCAAGAA 2775
Db 1001 AspAlaAlaIleLysIleLysArgTyrLysAspGlnAsnIle-----Gln 1016
OY 2776 GTTACTGACCCA-----AAACACAGACAACTATCTGTGAGCTATACGTACGT 2826
Db 1017 LeuThrAspProIleGlnGlnLysAspGlnLysAlaMetCysArgAlaValAlaTyrSer 1036
OY 2827 TTTCGAGATTAAGTATTAATTCGAGAAAGATCTCTGAGAAAGAAAGCGGATG 2886
Db 1037 PheAlaAspLeuGlyAspIleLysArgLysArgAspMetTyrAspGluAspLysSer 1056
OY 2887 GTAAGCTGACAGACATTTGGAACACTTTTGTATATACATATACATACGCAAGGC 2946
Db 1057 ThrAspMetGluThrArgLeuIleThrValPheLysAsnIleLysGluLysHisAspGly 1076
OY 2947 ---AAGGAAATGATTAATGATGAT-----GCCCAAAATTTTAAATG 2994
Db 1077 IleLysAspAsnProLysTyrThrGlyAspLysSerLysProAlaTyrLysLeu 1096
OY 2995 AGGGAATAATGTTGGAGGATATAGAGCAAGATGAGGAGGAGCAATGATGATATA 3054
Db 1097 ArgAlaAspTyrPheLysAlaAsnArgHisGlnAlaTyrArgAlaMetLysCys----- 1114
OY 3055 AAATATTTGAAGATTAATCGGACACACACAAACAGTATGTTATGCGATATAGT 3114
Db 1115 -----AlaThrLysGlyIleLeuLys----- 1121
OY 3115 GATCATACACATTTGATGATATATCCCAAAATTAAGATGATGATGACGAGTGGCA 3174
Db 1122 ProGluMetProValAspArgTyrIleProGlnArgLeuArgTyrMetThrGluTyrPhe 1141
OY 3175 GAATGACTAGTCAAGGTGCGAAGAAAGAGATGATTAAGTGAAGGAAAGTGTAAAG 3244
|||||

Db 1142 GluTyrTyrCysLysAlaGlnSerGlnGluTyrAspLysLeuLysIleCysAlaAsp 1161
OY 3235 TGTAAAGATTAAGATTAATGCTCAAGGCTGTACGAAAGACAGTGGTACAGTTAGCAG 3294
Db 1162 CysMetSerLysGlyAspGlyLys---CysThr---GlnGlyAspValAspCysGlyLys 1179
OY 3295 TGCACAGAACCTTGAATGAATATATGATTAATGATTAATGATTAATGAAAGCAATG 3354
Db 1180 CysLysAlaIleCysAspLysTyrLysGluGluIleGluLysTyrAsnGluGlnTyrArg 1199
OY 3355 ATATATACATTAATTAATCAAGATTTACATGACACACAAATGCTGTAGTAATAGT 3414
Db 1200 LysIleSerAspLysTyrAsnLeuLeuTyrLeuGlnAlaLysThrThrSerThrAspPro 1219
OY 3415 GGTATTAAGCTTCCAGTACTGCGCAAAATATCATATAGC-----AGCAATGTTAT 3465
Db 1220 Gly-----ArgThrValLeuGluLysAspAspAspProAspTyrGlnGlnMetVal 1235
OY 3466 GAATTTTGTGCGAATTAATACAAACAAATGTTGCGCAAA-----AGTAATMAA 3513
Db 1236 AspPheLeuThrProIleHisLysAlaSerIleAlaIleArgValLeuValLysArgAla 1255
OY 3514 AGTGTACTAGTATGAAGTGTGCTGATGCTGATGCTGATGCTGATGCTGATGCTGAT 3573
Db 1256 AlaGlySerProThrGluIleAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 1275
OY 3574 GCATATCTCCATGATACAGAAATTTGATGATGATGATGATGATGATGATGATGATGAT 3633
Db 1276 GlyTyrIleHisGlnGluIleGlyTyrGlyCysGlnGlnGlnGlnGlnGlnGlnGln 1295
OY 3634 GAAAA-----AGTATGATTAAGATTAAGCAAAATATGCTTTAGA 3675
Db 1296 LysLysHisGlyAlaThrSerThrSerThrThrLysLysLysLysLysLysLysLys 1315
OY 3676 GATTAACACAGACGACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3735
Db 1316 GlnProProProGluTyrAlaThrAlaCysAspLysIleAsnArgSerGlnThrGlu--- 1334
OY 3736 GTACAGATTAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 3795
Db 1335 -----GluProLysLysLysGluGluLysLysLysLysLysLysLysLysLysLys 1351
OY 3796 GATTAATTAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 3855
Db 1352 LysIleLeuGluGluLysAsnGlyArgThrThrValGlyLysCysAsnProLysGluSer 1371
OY 3856 AGTAATGATATCCGATGCGCAATGC---GAAATTAATTAAGTGAAGCAACCCGCT 3912
Db 1372 -----TyrProAspTyrAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1387
OY 3913 GTGTGATGCCCCCTAGACAAACAAAGTATGCTGATCAATTTCTGCAAAATGATGAAG 3972
Db 1388 AlaCysMetProProArgGlnLysLeuCysLeuSerTyrGluIleAlaHisGluSerGln 1407
OY 3973 ATAAAAAATTAACATGACAGTAAATTAATTAAGCACTTTCATCAATGTCGACGCA 4032
Db 1408 ThrGluAsnIleLysThrAspAspAsnLeuLysAspAlaPheIleLysThrAlaAlaAla 1427
OY 4033 GAAACATTTCTTCATGCTGATTAATTAAGAAAGTAAAG---GATGCTGAAGAAATGAACTC 4089
Db 1428 GluThrPheLeuSerTyrGlnTyrLysSerLysAsnAspSerGluAlaLysIleLeu 1447
OY 4090 GATTAAGATTAAGAAAGCAAAATTCCTCCGATTTTGTGAGATCCATGCTTACACA 4149
Db 1448 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetMetTyrThr 1463
OY 4150 TTGTGAGATTAATAGATTTTATTTTGAACAGATATATCAAAAGTCAATGATGAGGGA 4209
Db 1464 PheLysAspTyrArgAspIleCysLeuAsnThrAspIleSerLysLysGlnAsnAspVal 1483
OY 4210 AGTAATTAAGAAAGCAATGATGCTGCTTTTCAAAATGCTGACAAATGCTGCTTAAT 4269
Db 1484 AlaLysAlaLysAspLysIleGlyLysPhePheSerLysAspGlySerLysSerProSer 1503
|||||

QY 4270 GGAAGAACAGCCAGAGATGGTGCAGAGACATAGATGATGATGGAAGCTATGCTA 4329
Db 1504 Glysenserarglnglutriptylserthrasnlylproguilletpilsyglmetleu 1533
QY 4330 TGTGCTACTAGTAAAAATTTGGGCAAAAAAGATGAT-----TTTACCGAAACTAC 4380
Db 1524 CysAlaLeuThrLysTyrValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr 1543
QY 4381 GGTTCACACACGTCATTAATTAGTACAAA---AGCACCTTTGGAGATTTGCCAAA 4437
Db 1544 SerTyrAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnLuphAlaIa 1563
QY 4438 CGACCCAGATTTTACGATGGCTACCGAATGGTACGACGATTTGCTATACAGCAAA 4497
Db 1564 LysProGlnPheLeuArgTyrMetIleGlnTyrGlyLuphCysAlaGlnArgGln 1583
QY 4498 AATATTTGAAGATGGCGGAAAAATGTAATCA---AATGACCAATTGAAGTGTAT 4554
Db 1584 LysLysGlnAsnIleIleLysAspAlaCysAsnGlnLuphLeuSerThrGlnGlnCysAsn 1603
QY 4555 ACA-----GAATGTAATAAGAAATGCGAGACTACCTTAATATATGAAA---AAA 4602
Db 1604 AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGlnLuphLysVal 1623
QY 4603 AAAAAGAGTGATTCACAGATTAATATATACAGATGACGCGACAAAAAAGATTC 4662
Db 1624 LysLysGlnPheSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln 1643
QY 4663 GATGACACACATTTGGTGAATGTTTACAGACTATACGAAAGCAATGCAACAGTTTC 4722
Db 1644 AspProGlnTyrLysGlyTyrGlyTyrLysAsp----- 1654
QY 4723 TGTACACAGAAATTTACTCTAGTGTGTATAGCCTGGAAGTCCCTGTGTGACAA 4782
Db 1655 -----GlyValGlnPro-----IleGln 1660
QY 4783 AGAATATATCAATTTGTAGAAAAACAGGTTACTATGATCCGACAAATTTGGTGC 4842
Db 1661 GlyAsnGlnTyrLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1676
QY 4843 ACMAATTTATTTGMAAATGACGACAAATATACTAATTTGAGTAAAGTACAA 4902
Db 1676 ----- 1676
QY 4903 GGATTAAGTAAGGACCAACACAGGTCTATTAAAGTGCAAAACAAAGTCTAATAAC 4962
Db 1676 ----- 1676
QY 4963 TACAATTAAGTAAGTAAGTAAGTAAGTGTCTTTTCTCTGTGTGACTAGTATA 5022
Db 1676 ----- 1676
QY 5023 TGTTCATGCAATTTGATGGCAATTAACAGATCCAGAAATTAAAGTAAATGGTTG 5082
Db 1677 -----MetAspGlyAsnValLeuSerValSerProLysGlnLys----- 1689
QY 5083 CGAAAAAGATGATGAGAGTGGCGGACGACGAGGTACAAATTTGGTCAATACTACAA 5142
Db 1690 -----ProPheGlyLysTyr----- 1694
QY 5143 GAAAAAAGAAAAAGAAAAATATAAAGCTGCGATGCGCAAAATTTCTTATAGTGC 5202
Db 1695 -----AlaHisLysTyrPro----- 1699
QY 5203 CCGCTTGTAGTCTATGCAATATAGTTTATATGATTAAGATATATATCTAGATAT 5262
Db 1699 ----- 1699
QY 5263 GATATATTTGAAGATGAAAAACAAAGACCGAGAAAAATTTGAAGAAATATTTAAACAA 5322
Db 1699 ----- 1699
QY 5323 AATGGAATCACTTGTGCAAAAGAGTGAATAGTACTACAGAAAAATCCGTAAGTACGCG 5382
Db 1699 ----- 1699
QY 5383 CGAAAAATTTTCTGGAACGAAAAATAGAAATGTGTGGAACGCATGATATCGGGTAC 5442
Db 1699 ----- 1699
QY 5443 AAACGTGTAGGATGATGGAATAGTGAATAGTGAAGAAAGTGAAGATCTAATA 5502
Db 1699 ----- 1699
QY 5503 AATGTGTCTGTACTTCAATGATGATTAATCTATGAGGAAAAATCGCATGAAGT 5562
Db 1699 ----- 1699
QY 5563 ACTGCTATCAGTTTCTTCGATGTTTGGCGAATGGGGTGAAGATTTTGCAAACATAA 5622
Db 1699 ----- 1699
QY 5623 GAAAAGAATTTGAGAAATTTGTTAGGGCGCTGAATGATTAATCTTGTGTATATGAA 5682
Db 1699 ----- 1699
QY 5683 GATTAAGAAAGAAATGTACAGATGCGTGTACAAATATTAATAATTTATAGTAGTG 5742
Db 1700 -----GlnLysCys----- 1702
QY 5743 AAACCAAGTATGAAAAACAATCAAAATATATGTGAGAAATTAAGACAAATATATTC 5802
Db 1702 ----- 1702
QY 5803 GAGCATCCTGTGCGAAAAAGATGCAGAGACGCTGCGCAATATTTAGACAAACATTA 5862
Db 1702 ----- 1702
QY 5863 AAAATTTGTGAATAAAGTGAAGATTTGAAATATAGTATGAAAGATGTGCCA 5922
Db 1703 -----AspCys----- 1704
QY 5923 CAGCATTAAGTATGATTAATAGTCAAAATATATGCCCCATATTAGACGATGAACAA 5982
Db 1704 ----- 1704
QY 5983 GAAGTTGAAGAAAGTGAATTTGCAAGTGTGCCAGAGSTGCACACGCTACGAAGGAA 6042
Db 1705 -----TyrGlnGlnLys-----HisValProSerIleProPro----- 1715
QY 6043 ACACCGTCCACGAGGTATCACTGATATCAAAAGCAGCGCATCGAAAAAAGACGAAA 6102
Db 1715 ----- 1715
QY 6103 ACAGCGCGCTTACAAAAACAGCCGAAAAAAGTGAATAATTTAACACAGAAATCGGACA 6162
Db 1716 -----ProProProProValGlnProGln----- 1723
QY 6163 CAACACGAAACCCAGCAGCAGACAAACAAACGAAAGAACATCAACAGCAACA 6222
Db 1724 -----ProGlnAlaProThr----- 1728
QY 6223 ACAGATCTGACGTGGGCACAAATGTAAGGCCATTTCTTGATTAACCGCATAGCAGG 6282
Db 1729 ValThrValAspValCysSerIleValLysThrLeuPheLys-----AspThrAsn 1745
QY 6283 GGTGGAATGAGGGTGTATATCCAAAAAGCATATGAGCAATAT-----CCTAAATGGCGT 6336
Db 1746 AsnPheSerAspAlaCysGlyLeuLys---TyrGlyLysThrAlaProSerSerTyrLys 1764
QY 6337 TGAATTT-----GTAGTAAAGTCAATAAGAAAT 6363
Db 1765 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrLysSerGlySerAsp 1784
QY 6364 GAAAAATGCAATATGATGCTCTCTAGGACAAAAAATATATATATATATATATATAT 6423

```

Db 1785 SerGlySerIleCysIleProPheArgArgLeuValGlyLysLeuIn 1803
QY 6424 TTAATAATGAACT-----GAAATAAGCGTGAC 6453
Db 1804 ----GluTrpAlaThrAlaLeuProGlnGlyLysAlaAlaProSerHisSerHisAla 1822
QY 6454 AATGATATTAAGAAGCGCTTTAATTAATGTGACAGCAATAGAACTCAATTTTGTGTTA 6513
Db 1823 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuTrpAsp 1842
QY 6514 AATATATATATGAA--AATCTGCAGACGAAATGAA----- 6549
Db 1843 ArgTyrLysGluGluLysLysProGlnGlyAspGlySerGlnAlaLeuSerGlnLeu 1862
QY 6550 -----TTGCAAAAATGCAACA 6564
Db 1863 ThrSerThrTyrSerAspAspGluGluAspProProAspLysLeuGlnAsnGlyLys 1882
QY 6565 ATTCGCAATGCAATTTAAAGCAATATGATATATACATATGCTGATTTAAAGATATGTTT 6624
Db 1883 IleProProAspPheLeuArgLeuMetPheTyrThrLeuGlyAspTyrArgAspIleLeu 1902
QY 6625 TTGGAACTGATATTTCTATGATTAATAAATAATATACCTGTAACCAATAGTGTACACACC 6684
Db 1903 ValHisGlyLysAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1922
QY 6685 ATTCTCAATGCAATTAATAAGCAATATGATATATACATATGCTGATTTAAAGATATGTTT 6744
Db 1923 ValLeuGlnAlaSerGlyAsnLysGluAspMetGlnLysIleGlnGluLysIleGlnGln 1942
QY 6745 ATA----- 6747
Db 1943 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1962
QY 6748 ----TTTGGAGAAATAATAAATTTATTTGGAGAGGATGATATGATTAACCTAT 6804
Db 1963 LysTrpAsnGlnHisAlaGluSerIleTrpLysGlyMetIleCysAlaLeuThrTyr 1982
QY 6805 -----CATCTCACAGACGAAAGCAAAAAAGAAATAATAGA 6840
Db 1983 ThrGluLysAsnProAspThrSerAlaArgLysAspGluAsnLysIleGluLysAspAsp 2002
QY 6841 GATATATAC----- 6849
Db 2003 GluValTyrGluLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2022
QY 6850 -----CAGTACAAAT--GACATGACCAAACTG----- 6873
Db 2023 ThrGlyThrTyrTrpThrGlnTyrAspTyrGluLysValLysLeuGluAspThrSerGly 2042
QY 6874 -----ACGCTTCCCTGGAAGAGTTTGTAAAAAGGCC 6906
Db 2043 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheValLeuArgPro 2062
QY 6907 CAATTTTGGATGCGTTACACAGAAATGGGCAAGAAATTTCTATTAAGAGAGAAAGAACAG 6966
Db 2063 ProTyrPheArgTyrLeuGlnGluTyrPglLysAsnPheCysLysLysArgLysHisLys 2082
QY 6967 TTGTTAAATTTGAGCGGCGCTGTAGGAATATAGAG----- 7002
Db 2083 LeuAlaGlnIleLysHisGluCysLysValGluGluAsnGlyGlyLysSerAlaArgGly 2102
QY 7003 -----TGTAAAT-----GGTAGTAAT 7017
Db 2103 GlyIleThrArgGlnTyrSerGlyAspGlyGluAlaCysAsnGluMetLeuProLysAsn 2122
QY 7018 GAGGATAGACACAGAA-----TGTCAAGAGCGGTGTGTAACATATCAA 7062
Db 2123 AspGlyThrValProAspLeuGlnLysProSerCysAlaLysProCysSerTyrArg 2142
QY 7063 AATTTTATTAAGAAGTGAATACTGATATGAAGACAAAGAGAAAGTCAAAAGAT 7122

```

```

Db 2143 LysTrpIleGluSerLysGlyLysGluPheGluLysGlnGluLysAlaTyrGluGln 2162
QY 7123 AAAGAT-----GGCAAAAAGATATAGCAATATCTTCTACTGAAAGACATA 7170
Db 2163 LysAspLysCysValAsnGlySerAsnLysHisAsp----- 2174
QY 7171 GAGNAGGCAACATGTGCTCATGAATATTTAAACATGAAATTAAGAAATTAATGTGCAAT 7230
Db 2175 -----Asn 2175
QY 7231 AAGGATGTCTTGTATGCAAAAAACCTTTCAACACTACCAAAAAACAAACAATCA 7290
Db 2176 GlyPheCysGluThrLeu-----Thr 2182
QY 7291 CAATCATCCGATGCTATGATATATGCGAGAAATCGCTGATATGTCTCGAAGATTTAAC 7350
Db 2183 ThrSerLysAlaLysAspPheLeuLysThrLeuGly----- 2195
QY 7351 AAGTGTGAGTGTCTGCAACTTCAAAAAAGGATCTATGATT----- 7392
Db 2196 ----ProCysLysProAsnAsnValGluGlyLysThrIlePheAspAspLysThr 2213
QY 7393 -----CATCAAAAAAATTTACTGAACT-----AAATACCTATGATGTGTAGAG 7440
Db 2214 PheLysHisThrLysAsp---CysAspProCysLeuLysPheSerValAsnCys----- 2230
QY 7441 AAAGCAGCATATTTATTTATCTAAAGACAGCAAAAAATTAATGATATTTACTTGAAGAA 7500
Db 2230 ----- 2230
QY 7501 AATTTATACCTATTTAGTGTACAAAGCAAAAGCAATTAATAGTGTGACTAATAT 7560
Db 2231 -----LysLysAspGluCysAspAsnSer----- 2238
QY 7561 AATCCTTGCGATCCTTAGAAACCTTATGCAACCTGATTAATATATAGAGAGAAACCT 7620
Db 2239 -----LysGlyThrAsp 2242
QY 7621 TGTGAAATATAGAGAAATAATCGTTTAAAGTAGATTATGAAATGAAATGTTCACAAAT 7680
Db 2243 CysArgAsn-----LysAsn 2247
QY 7681 TCAAGTTCTATCAGAGAAAAAAGAGTATGTGTACTCTCCAGAAAGAAACATATGTGC 7740
Db 2248 Ser----- 2248
QY 7741 TTAAGCAATTTAGATGCAATTAATAATTTGAAGAAGCTTAAGATAGTAATATCTCTTAAA 7800
Db 2249 -----IleAspAlaThrAspIleGluAsnGlyValAspSerThrValLeuGluMet 2265
QY 7801 ATGGTTCGTGCAACTGCAGCAAAATGAAGAAATGACATTAATAAACCCTCAACTCA-- 7857
Db 2266 ArgValSerAlaAspSerLys-----GlyPheAsnGlyAsp 2278
QY 7858 -----GAGAACGGGTGCGCAATGCAATCCATATATGTATACTATGAATATAGTTGCT 7911
Db 2279 GlyLeuGluAsnAlaCys----- 2284
QY 7912 GATCTGGGTGACATAGTATAGAGAAACAGATATGTACGAATTTGGTGTACTTACTCC 7971
Db 2285 -----ArgGlyAlaGly----- 2288
QY 7972 GTAGAAATTAATAATTAATTAAGGTTTGAATACATATATGAAATAAGAGAAATTAATAAT 8031
Db 2289 -----IlePheGluGlyLe----- 2293
QY 8032 AAAGGTAGAAATTAATTAACACGATGTACAAACGTTTCTGTTCTGTGTGGATGCTAAT 8091
Db 2293 ----- 2293
QY 8092 AGAAAGATATTTGAAAGCA-----ATGACGTGCAAGACAGCAAA 8133
Db 2294 ArgLysAspGluTyrLysCysArgAsnValCysGlyTyrValValCysLys---ProGlu 2312

```



```

Db 375 ArgPheValIysTrpIleAspAsnGlnArgLysGlnPheAspLysGlnLysLysTyr 394
OY 1210 GAAAAAGAAAACAA-----TCATATTATGCAACGATTAACAAATTTGTCAT 1257
Db 395 AspIleGlnIleAsnLysThrHisGlyThrThrIleThrGlnAsnGlyLysIleAsn 414
OY 1258 AATATTAACTGAAATATTAAACAAATTTATGAAAAAAGTTAGAAAGCAATATGCA 1317
Db 415 AsnLeu-----TyrValGlyHisPheTyrLysIleLeuLysLys-----TyrTyrPro 430
OY 1318 ACTAATGACACTTTTAAATTTTCTAAATGAGAAAGAAAGTATGTTAA-----GGAGA 1371
Db 431 ThrValAspLysSerLeuGlnLysLeuAsnAspGlnLalIleCysLysLysProProAsn 450
OY 1372 TTACACAGAGAAAAGATATTACTTTTACACAGCTGCTGATGCAAGGATATTAT 1431
Db 451 ValGlyAsnGlnLysAlaSerThrValAspPheAsnAsnGlnLValAsnThrThrPheSer 470
OY 1432 CGTTGAGAAATTTGCCAAGTGTGTCGCCGACTGCGGGGTCAATGTGAT----- 1479
Db 471 HisThrThrTyrCysGlnLalCysProTyrCysGlyValGlnLysGlnLysAsnGlyGly 490
OY 1480 GGTATATAATACACACACAAATCA---GATAATGATCGTGAGCTGTAATATATGAAAC 1536
Db 491 GlyTrpLysAlaLysGlnLysSerCysAlaLysLysGlnLalGlyIlePheAsnLysGln 510
OY 1537 TATAAACCTCCACGAGGGGTGAAGCTACTAATATCATCTGCTTATAGTGTAAAGAA 1596
Db 511 -----AsnSerThrAspIleLysIleLeuThrProGlnLysGly 523
OY 1597 CAGGTGATATTACAAAAATATAGAAAAATTTTGTAAACGCTCAACTAATTACAAAGAT 1656
Db 524 ArgSerLysThrLeuLysLysLysLysThrPheCysLysAspLysGlnLysIleLys--- 542
OY 1657 AAAAATATCAAAAAATGGAAATGCTATTATTAAGATGAATGAATAATATGATGTAACAG 1716
Db 543 -----AsnAspIleTyrLysCysHisTyrAspAspAsnGly-----Thr 555
OY 1717 GAACAAAATATCTGCAATCAATAATGAT----- 1743
Db 556 AspAspGlnThrAspAspSerAsnAspCysValLeuGlyAspTrpLysLeuThrLys 575
OY 1744 AATCTTAATATATATCATTTATATTTTGAATTTGGTATGATATTTATTAAG 1803
Db 576 GluAspLysIleLeuSerTyrAsnAlaPheThrPheTyrValHisAspLeuIle 595
OY 1804 GATCTATTATAGTGAATGACAAACTTAAACTGTATTAAT---AATCAACACGACAT 1860
Db 596 AspSerIleLysTrpArgAspGlnHisGlyArgCysIleAsnLysAspLysGlyLysThr 615
OY 1861 TGTATTGATGAAATGTAACGAAATGCTAATGTTTGGACAGATGGGTTAAACAAAAGAA 1920
Db 616 CysIleLysGlyCysAsnLysLysCysIleCysPheGlnLysTrpValGlnLysLys 635
OY 1921 GAAGATGAAATAGTATTAAGAAAGAACTGTTCACAAAAAAGAAATATACAGACATGAT 1980
Db 636 ThrGluTrpGlyLysIleLysAspHisPheArgLysGlnLysAspIleProLysAspTrp 655
OY 1981 TATATATATTAATTAATCTTTTGAAGGTATATTTTAAAGTTATGATTAACCTGAC 2040
Db 655 ----- 655
OY 2041 AAAGATGAAGCAAAATGGAAGAACTTATGCAAAATATAAAAAGAAAAAATGACGTTT 2100
Db 655 ----- 655
OY 2101 TCCAAATTTGAAAATATATAGGACTATTATAGATGCAATA-----GAAGCTCTGTTA 2154
Db 656 -----ThrHisAspAspPheLeuGlnThrLeuLeuMetLysAspLeuLeu 671
OY 2155 GATCACTTAAAGAAACCTGCCACGATATGTAAGCAATAATATACAAAGAAAGCATGTGAA 2214
Db 2214 ----- 2214

Db 672 GluIleIleGlnAspThrTyr-----GlyAspAlaAsnGln----- 683
OY 2215 ACATCCCATTAATATCAACAAACAAACCCGTTGTTAAACCTGTGAGGACGCAACCCACT 2274
Db 683 ----- 683
OY 2275 AAAAATATTAAGAAATATGACAAATCTTAAAGAGATGATACGAGACGCAAGAAAT 2334
Db 684 -----IleLysArgIle-----GluAlaLeuLeuGlnLAla----- 694
OY 2335 CGTGTCTTCATTAATTTGAAGAAAGCAGACGACGAGATATATTAACGTGGGGGTAG 2394
Db 695 ---GlyValGlyGlyLysAspPheAlaIleAlaGlyLeuTyrThrLysGlyPheVal 713
OY 2395 AGAAAGACTTCAAGGACAAATTTATGATTAATGATTAATGATTAATGATTAATGAT 2454
Db 714 AlaGlu-----LysAspThrThrIleAspLysLeuLeuGlnHisGlnLys----- 729
OY 2455 GGTTTTCAATATGACCAATGATGTGCAAGGACGAGTGTATGATCAACAAAGATTT 2514
Db 729 ----- 729
OY 2515 GTCTGAGAACTGAATGGAGAGTGCATCCGAAACATGCTTAAGATCAGCAAGATGT 2574
Db 730 -----GluAlaAspLysCysLeuLysThrHisThrAsp 741
OY 2575 ATTATGCTCTCCTAGAAGACGACATATATGTACATCAATTTGGAACATTTCAACAGGAT 2634
Db 742 ThrCysProProGln-----Glu 747
OY 2635 GATCACCCCACTTAATGTAATATGTTGATGATTTAGTTAATATCTTTTGGGGAT 2694
Db 748 AspArgSerVal----- 751
OY 2695 GTTCTTATATCAGCAAAATATGACAAACAAAGATTAATACGATGTATTAAGAAAGAT 2754
Db 752 -----AlaArgSerGlnLysSerAlaThrVal----- 759
OY 2755 AACTAAAGGGCCCAAGAAAGTAACTGACCCCAAAACACGACACATCTGTGACGCT 2814
Db 760 -----ProSerProAlaAspProLysAlaThrGlnGluVal----- 772
OY 2815 ATACGTACAGTTTTCAGATATAGTGATTAATATTCGAGAAAGATCTGTGGAAAGA 2874
Db 772 ----- 772
OY 2875 AACGGTGACATGTAAACCTGCAAGGACATTTGCAAACTGTTTGGTAAATATACATPAAG 2934
Db 772 ----- 772
OY 2935 TCACTAAGAGCAAGAAATGATTAATATATATGATGATGCGCCCAATATTTAAATTTG 2994
Db 772 ----- 772
OY 2995 AGGAAATTTGTGGAAAGCTAATATAGACCAAGATATGGAAAGCATGAATGTATATA 3054
Db 773 -----AspAlaAsnAlaSer----- 777
OY 3055 AATATTGGAAGATTAATGCGGACACCAATCAACACAAAGTAGTTATTCGCGATATAGT 3114
Db 778 -----Ser 778
OY 3115 GATCATCACCATTTGATGATATATATCCCAAAATATTAAGATGATGACCGAATGGCA 3174
Db 779 AspAspGln-----AspAspPheGlnGln----- 787
OY 3175 GAATGTACTTCGAAGGTGCGAAAAAGAGATATGATTAAGAGGAAGATTAAGAG 3234
Db 788 -----GluGlnGlnGlnGlnLysLysLysLysLysLysLysLysLysLysLys 801
OY 3235 TGTAGATTAAGATTAATGCTCAAGGCTGTAGGAA-----GAGAGTGTACAGGT 3285
Db 802 ValGlnGlnLysThrAspGlnSerAlaThrGlnLysLysLysLysLysLysLysLysLys 821
```



```
QY 3286 TGTACAGATGACACAGAGCTTGTATGAATATATGATATATAGATATATGAGAAAGAA 3345
    |||:::
Db 822 ThrThrGln----- 824

QY 3346 CAATGGAATATATATACAGATTAATACAAAGATTTACATGACACAGACAAATCTCTGT 3405
    |||:::
Db 825 -----AspGlyValLysProAlaSerGlnGluAspValLysVal 838

QY 3406 AGTAATAGTGTATGTAAGCTCCAGTACGCCAAAATGATATAGACAGAAATGTATT 3465
    |||:::
Db 839 CysSerIle-----ValAspLysAlaLeuLys 847

QY 3466 GAATTTTTCGCAATTTATCCACAAATGTCGGCAAAAGTATTAAGTGTAGT 3525
    |||:::
Db 848 GlyLysLeuAspAspAlaCysThrLeuLysGlyLysThr-----AlaProThrSer 865

QY 3526 GATGAAGAAGCTCTCATTTGTACTAACACACAGATGGAATGTTGAGAGATATCTCAT 3585
    |||:::
Db 866 TrpLysCysIleProSerGlyAsnAsnThrThrGlnSerThr----- 880

QY 3586 GATACAGAAATTTTGTATGATGTCAAGTACAAAATGATTTGTATGAAAAAGTAT 3645
    |||:::
Db 880 ----- 880

QY 3646 GGTAGAGATACGAAAAATATGCTTTAGAGATTAACACAGACCATGATGCTGCTGT 3705
    |||:::
Db 881 -----ThyLysPro-----GlyAlaAla 886

QY 3706 GGTGTTAAAGTGGATGGAACGACAAAGGTCACAGATTAACAAAAACGAAACGGGAA 3765
    |||:::
Db 887 GlyThrProSerGly----- 891

QY 3766 GAAAGCATGACGGAATGTAACAGTGAATGATATCTTAAGAAAAAGTGAAGAAA 3825
    |||:::
Db 892 ---LysAspThr----- 894

QY 3826 CAAGTACAGATTTGTATCCAAAAAGAAATGATATGATCCCGATTGGCAATGCGGA 3885
    |||:::
Db 895 -----Gly 895

QY 3886 AATATTAATTTAGTGAAGACCGCTGCTGTGATGCCCTAGAACACAAAGTTATGTC 3945
    |||:::
Db 896 SerIle-----CysValProProAlaGatGatLysLeuLys 907

QY 3946 GTACATTTCTTG----- 3957
    |||:::
Db 908 ValGlyLysLeuHisAspTrpAlaGlyGlyLysThrGlnAlaLysSerGlnGlyThr 927

QY 3958 -----GCAATGATATAT----- 3969
    |||:::
Db 928 SerGlyLysLysThrProSerGlyAsnGlnLysSerProSerGlyLysLeuProGln 947

QY 3970 -----GAAATTAATAAATTTACATGACAAAGTATTAATAAAGAAAGCTTATC 4017
    |||:::
Db 948 GlyProThrProGlnThrLysGlnThrProLysSerLeuHisAlaAlaVal 967

QY 4018 AATATGTCACACAGCAAGAAATCTCTCATGATATTAATAAAGTATTAAG----- 4068
    |||:::
Db 968 SerProProAlaGluAlaGatGatGatGatGatGatGatGatGatGatGatGatGat 987

QY 4069 -----GATGTTGAAGCA 4080
    |||:::
Db 988 GlnHisGlyAlaGlyAlaThrGlyGlnGlnThrLysLeuLysLeuAspGlyGly 1007

QY 4081 AATGAAGTCTC-----GATTAAGAAATTAAGAAAGCAAAATGCTCCGATTTGAGATCC 4137
    |||:::
Db 1008 GlnGlnThrProAspLysLeuLysThrGlyHisIleProProAspPheLeuAlaGln 1027

QY 4138 ATGTTTACACATTTGAGATATAGAGATTTTATTTATTTGA-----ACAGATATA----- 4188
    |||:::
Db 1028 MetPheLysThrLysLeuGlyAspTrpArgAspIleLeuValGlyAsnThrAspIleVal 1047

QY 4189 -----TCAAAAGCTATGCTGAGGCAATTAACCTAAAGACCAAAATGATCTT----- 4236
    |||:::
Db 1048 HisThrSerGlyAsnLysGluAspMetGlnIleMetGlnAlaIleGlnLysLysIleGlu 1067

QY 4237 -----CTTTCAAAAATGAGTACCAAAAATTCCTTATGGAAGAAACA----- 4278
    |||:::
Db 1068 GlnIleLeuProThrSerGlySerSerProSerProProAlaValThrGlnThrGlnHis 1087

QY 4279 -----CGCCAAAGATGCTGACACAGAAATGATGATATGATATGGAAGT 4333
    |||:::
Db 1088 SerValGlnAsnProAlaGlyLysThrTrpAsnGlnLysLysIleTrpGlnGly 1107

QY 4324 ATGCTATGTCACATGTA----- 4341
    |||:::
Db 1108 MetValCysAlaLeuLeuThrTrpAsnThrAspThrProSerGlyThrAlaProThrGlnIle 1127

QY 4342 -----AAATTTGGGCAAAA-----AAGATGATTTTCCGAAAC-----TAC 4380
    |||:::
Db 1128 GlnGlnValAlaThrThrLysLeuAlaGaspLysAsnSerLysAsnProLysIleProGlnTrp 1147

QY 4381 GGTACAAACAACGCTCAATTTAGTGAC----- 4407
    |||:::
Db 1148 LysTrpAspGlnValLysLeuAspAspThrSerAspAlaLysThrThrGlySerProVal 1167

QY 4408 -----AAAGCACCACTTTGGAGAGATTTGCCAAAGCAGCCAGTTTATGCA 4455
    |||:::
Db 1168 ProSerGlyLysLysIleThrProLeuThrAspPheIleSerArgProProLysPheArg 1187

QY 4456 TGGCTAACCGAATGTCAGCAGTATGCTATACAGCAAAAATTTGAAGAGATGTG 4515
    |||:::
Db 1188 TyrLeuGlnGlnTrpGlyGlnThrPheCysLysGlnGlyLysAspGlyLysIle 1207

QY 4516 CAGCAAAAATGTATGTA----- 4533
    |||:::
Db 1208 LysGlnGlnLysAspArgLysAspArgThrGlyHisGlnHisCysSerGlyAspGlyLysAsp 1227

QY 4534 -----AATGACCA-----TTGAAGTGTATGACGAA 4560
    |||:::
Db 1228 CysThrThrThrAspAlaAspArgAsnAspLysPheValAspLeuAsnAspCys---ArgAsp 1246

QY 4561 TGTATTAAGAAATGCGAGACTACGTTAATATATGAAA-----AAAAAAAAGATGATG 4617
    |||:::
Db 1247 CysHisIleGlnCysAspGlyLysTrpArgLysTrpLysLysLysLysLysLysLysLys 1266

QY 4618 CCACAAATTAATTTACAAAGATGAACGCGCAAAAAGATGATGATGACAAACACAT 4677
    |||:::
Db 1267 LysGlnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1283

QY 4678 GGTGTAATGTTACAGACTTACTGAAACGAAATGCAACAGATTACTTGAACAGGAAATTT 4737
    |||:::
Db 1283 ----- 1283

QY 4738 ACTGCTAGTGTGCTGATTAAGCCTGGAAGTGCCTCTGTGATCAAAAGAAATATGCAATG 4797
    |||:::
Db 1284 ---SerSerGlyLysAsp----- 1288

QY 4798 TTACAAAACAGGCTTACTATGATCCGACAAACATTTGGTGCACAAAATTTATTTGAA 4857
    |||:::
Db 1289 ---AsnAsnCysCysLysAspIleGlnLysHisLysSerAlaAlaValPheLeuLys 1306

QY 4858 AATGACGACAAATATACTAATCATTTGCAAGTAAGATAGTCAAAAGATTAAGTAAAGAG 4917
    |||:::
Db 1307 Gln-----LeuLysHis 1310

QY 4918 GCAAAACAGGCTGCTATTAAGTGAACAAACAAAGCTCTAATATACATTAATCTGAAA 4977
    |||:::
Db 1311 CysLysAsnGlnGlnThrSer-----GlnAsnLysGly---AsnGlnGlnAspGlnLeuAsn 1338

QY 4978 GAATTTGACTGAAGATGCTTTT-----TCCCTTCCTGCACTGATATATGTTTTCATGC 5033
    |||:::
Db 1329 LysLeuAspPheAspLysIleProGlnThrPheSerProSerThrTrpCysLysAlaCys 1348

QY 5034 ATTGATGCGCAATTTATACAGATCCAGAAAGTTAAAGATGAAGAAATGGTTCGAAAAAGATT 5093
    |||:::
```

```
Db 1349 -----ProVal 1350
QY 5094 GATGGAGTGGCGGCAAGAGGTACAAVTTGGTCATACATACAAAGAAAAAGAA 5133
    |||
Db 1351 TyrglyValAsnCysAsnGly-----AsnLysArg 1360
    |||
QY 5154 AAAAAGAAAAATAAAAACGTGGATGCGCAAAATATTCTTATGAGTCCGCGCTGTAG 5213
    |||
Db 1361 GlyArgGlyGlyThrAsn---GlyCysThrThr----- 1370
    |||
QY 5214 TGCTATGAATATAGTTTATGATTTTAAAGATATATATCTAGTATGTAATTGGA 5273
    |||
Db 1371 -----AsnAsnGlu-ProGluAsnLysGly 1378
    |||
QY 5274 AGATGAAAAACAAAAGACCGAGAAATTTGCAGAAAAATTTTAACAAAAATGGAACATC 5333
    |||
Db 1378 uAsnAsp----- 1380
    |||
QY 5334 AGTTGGCAAGAGTGAATGACTACTACAGA-----AATCCGGTACTGCGCG 5384
    |||
Db 1381 -----LysGlyValAlaIleSerThrIleSerIleLeuIleAsnAspLysSerThr----- 1396
    |||
QY 5385 AAAATTTTCTGGAACGAAAAATAGAATGTGTGTGAGCAATGATATGCGGGTACAA 5444
    |||
Db 1396 ----- 1396
    |||
QY 5445 ACGTGTAAGGATGATGGAATAGTGAATAGTGCAGAAGTGAAGATCTAAAAA 5504
    |||
Db 1397 -AsnGlyAlaIleThrAsnGlyThrThrGlyThr-----ThrAspGluThrLeuLysGly 1413
    |||
QY 5505 ATGTGTTCTGTACTCTCAAGATGATGATATCTATG-----GGGAAAAATCGCATGA 5558
    |||
Db 1413 uCys-----SerAspLysTyrrAlaIlePheIleLysGlyLeuArgLysIle 1428
    |||
QY 5559 AGGTACTGCGTATCACTTCTTGATGATGTTGCGCAATGGGTGAAGATTTTGCAAACA 5618
    |||
Db 1428 uTrpThrCysGlnLys-----LysTyrrGlyValAsnGlnCys----- 1440
    |||
QY 5619 TAAAGAAAAAGAAATGAGAAATTTGTAGGGGCGGTATGATGATATGATCTTGCGATGA 5678
    |||
Db 1441 -----AsnLeuThrAsnArgValAsnAsp---ThyTrpPheAspLys 1453
    |||
QY 5679 TGAAGTAAAGAAAGAAATGTCACAGTCCGTGCACAAATATTAATTTATAGTGA 5738
    |||
Db 1453 sAsp-----IleValPheAsnGluPheGlnArg 1463
    |||
QY 5739 GTGGAACACACAGTATGAAAAAATCAAAAATATGATGGAATTAAGCAAAATATA 5798
    |||
Db 1463 gTrpLeuAlaGlyTrpHe-----ValHisAspTyrrAsnIleLeuLysHisLysIle-- 1479
    |||
QY 5799 TTCGAGCATCTCTGTGCGAAAAAGATGCAGAGCGCTCGCAATATTTAGCAAAACAAAT 5858
    |||
Db 1480 -----AspProCysIle 1483
    |||
QY 5859 AAAAAAATTTGTGAATATTAAGAGATTTGTCAATATATAGTATGAAAGATGTGTC 5918
    |||
Db 1483 eLysLysGluLysGlnAspLysThr-----GluHisLysCysIle----- 1496
    |||
QY 5919 CACACAGGATTAAGTGAATAGTCAAAATATGCGCCCATCATTAAGACATGAAC 5978
    |||
Db 1497 -----AsnGlyCysAsnIle----- 1501
    |||
QY 5979 AAAAGAAAGTGAAGAAAGTGAATTTGTCAAGTCCACAGAGTCCACCACGTGACGAAG 6038
    |||
Db 1502 -----LysCysGlnCys-----ValArg-- 1507
    |||
QY 6039 GAAACACCGTACACAGCGGTATCACTGATTCAAAAGCAGCGCATCGAAAAAAGAAC 6098
    |||
Db 1507 ----- 1507
    |||
QY 6099 GAAAAACAGCGCGCTACAAAACAGCGAAAAAAGTGAATAATCTAACACAGAAATGCG 6158
    |||
Db 1507 ----- 1507
    |||
QY 6159 AGCACAACACGAACCCGACGACGACACACAAACAGAAAAAGCAATCAACAGCAAC 6218
    |||
Db 1507 ----- 1507
    |||
QY 6219 AACACAGAAATCTGACGTGGGCAATAGTGAAGGCCATCTTGTGAAATTAACAGATAG 6278
    |||
Db 1507 ----- 1507
    |||
QY 6279 CAGGGGTGAATAGAGGTTGTAATCCAAAAACGATGACAAATATCTTAATGGGGTTG 6338
    |||
Db 1508 -----LysTrpLeuGly 1511
    |||
QY 6339 TATTGATGTAAGTCTAAAGAAAAATGCAATGATGATGCTCTAGAGAAAAA 6398
    |||
Db 1511 uLysLysGly-----AsnGluTrpLys----- 1518
    |||
QY 6399 ATTATGTAATTAATTAATCAATATTTAATTAATGAAACTGAAATTAAGCTGACATGA 6458
    |||
Db 1519 -----AsnLysLys-----LysHisTyrrAsnIleAsnSerAsnAs 1530
    |||
QY 6459 TATTAAGAGGCTTTTAAATTAATGTCAGCAATAGAACTCAATTTTGTGTTAAATA 6518
    |||
Db 1530 PAspLysGluThr----- 1534
    |||
QY 6519 TATTAATGAAATCTCTGACAGCAAAATGAATTCGAAAAATGGAACAATCCAGATGAAT 6578
    |||
Db 1534 ----- 1534
    |||
QY 6579 TAAAAAATTAATGTAATTAATATATGATGATTAATAGATATG---TTTTTGGACTGA 6635
    |||
Db 1535 -----IleAlaTyrrAsnValLysSerTyrrPheValAspGlnGlyLeuPheAspThrAs 1552
    |||
QY 6636 TATTCTTAATGATTAATAAATTAATTAATCTGTAACAAATAGTGAACAACATTCATGA 6695
    |||
Db 1552 PTrpLysLysAlaGlnLysValVal----- 1560
    |||
QY 6696 AAATTAATGAAGAAAAACAGATAAAAAAAGATGAGAAATTAAGTAAATTTTGGGA 6755
    |||
Db 1561 -----GluAspGluLysGluArgLysLysIleTrp-- 1570
    |||
QY 6756 GAAAAATTAATAATTTATTTTGGAGGAATGATATATGATTAATCTATCATCTACACA 6815
    |||
Db 1571 -----GlyCysThrGlyHis-----As 1576
    |||
QY 6816 CGAA---AACGAAAAAGAAAAAATTAAGATTAATTAACAGTAAATGACATGACCAACT 6872
    |||
Db 1576 pLysCysSerGluLysGluLysGluLysLysLysLysLysLysLysLysLysLysLysLys 1593
    |||
QY 6873 GACCCCTTCCTTGAAGAGTTTGAAGAAAAAGCCCAATTTTGAAGATGTTTCACAGAAATG 6932
    |||
Db 1593 uLysSerGluLeuGlnAspLysIleThrSerCysGln----- 1605
    |||
QY 6933 GCGCAAGAAATTTTGAATTAAGAGGAAGAAAGAGCTTTGAATTTGGAGCGGCTGTAA 6992
    |||
Db 1606 -----AsnLysHis----- 1608
    |||
QY 6993 GGAATATGACTTAATGTAATGTAATGACGTAAGACACAAGAAATGTCAGAGCGGTGT 7052
    |||
Db 1609 -----AsnProAsnGlyLysThr-----AlaCys-- 1616
    |||
QY 7053 AACATATCAAAATTTTATTAAGAGTGAAGAACTGAATGAAGACAAAGAGAAAGTT 7112
    |||
Db 1616 ----- 1616
    |||
QY 7113 CAAAAAGATTAAGATGCAAAAAAGTATAAGATTAATCTCTACTGAAGAGACATAGA 7172
    |||
Db 1616 ----- 1616
    |||
QY 7173 GAAGCAACATGTGCTCATGAATATTTAAACATGAATTTAAAGAAATTAATGTGCAATAA 7232
    |||
Db 1616 ----- 1616
    |||
```

QY 7233 GGATGTTCTTGTATGCAAAACCTTCTTCAACTACCAAAAAACACACATCACA 7232
Db 1617 -----AspProPheProSerProThrProGluGluTh 1627
QY 7293 ATCATCCGATGCTATGATATGACGAGAAATCGTGATTAATGCTCTGAGAAATTAACA 7352
Db 1627 rAspProLeuAspAspThrProAspProLeuAspAspAspGlnHisThrGluGlnPr 1647
QY 7353 GTGTGAGTGTCTGCAACTTTCAAAAAGGAGATCTATGATCATCAAAAAAATTACTGA 7412
Db 1647 olyPheCysPro-----Pr 1652
QY 7413 ACCTAAATATACCTATGATTTGTAGAGAAAGCATATTATTATCTAAAGACAGA 7472
Db 1652 oProProProMetThrCysValGluLysIleAlaLysGluLeuArgValGluAlaG1 1672
QY 7473 AAAATATATGATATCTACCTGGAAGGAAAA----- 7503
Db 1672 uGlyLysIleAsnAsnGluLeuLysGlyAsnGlyLysAspPheAsnGlyLysCysAsnAs 1692
QY 7504 -----TTTATACCTATTGAGCTCTCAAG--GAAAGAGAAAG 7538
Db 1692 nValLysLysAsnGlyAlaValIleGlyGluGluSerCysLysPheGluGlnThrTy 1712
QY 7339 TAAATATGCTTGACATAATTAATCTTCGATCTTAAGAACCTTATGCACTGATTA 7598
Db 1712 rGluAsnSerValAsnAsnIleAsn----- 1720
QY 7599 ATATATGAGAAAGAAACCCCTGTGAAATATAGAGAAATATGTTTAGAGATTA 7658
Db 1721 -----AsnLysCysLysAspAsnGlnAsnGluLysPheLysIleGlyG1 1735
QY 7659 TGAATGCAATGTTACAAAATTCAAAGTCTATCAGAG--AAAAAGAGATGTGT 7715
Db 1735 nLysTy-----AsnPheLysTyIleGlyThrIleArgLysAspLeuGlyC1 1751
QY 7716 ACCTCCAGAAAGAGACATATGCTTAAGGAATTTAGATTAATTAATTAATGAAGACT 7775
Db 1751 eProProArgArgLysLysMetCysLeuAspAspLeuSerMetLeuGlyArgThrTh1 1771
QY 7776 TAAAGATGATATATTCCTTAAATGATGTCGCAACCTACAGAAATGAAGAAATGA 7835
Db 1771 eSerAspSerSerAlaLeuLeuLysLysIleGlnGluAlaAlaLysSerGluAspAs 1791
QY 7836 CATATATAAAATCTCAACCTCAGAGAAAGCGGTGCAATGATCAATATGATACTAT 7895
Db 1791 pIleIleArgLysLeuLeuGlnGlnAsnSerCysAspGlnHisArgIleCysAspAlaMe 1811
QY 7896 GAAATATATGTTGCTGATCTGGGTGACATAGTATAGAGAAACAGATATGTTACGAATTGG 7955
Db 1811 tLysTySerPheAlaAspLeuGlyAspIleIleArgGlyArgAspLeuTrpAsnLysAs 1831
QY 7956 TGGTTACTTACCTCCCGCAAAATATAATATATAGTTTGAATCATATATGAAA 8015
Db 1831 nSerLysGlnLysGlyLeuGlnLysArgLeuGluLysAlaPheIleAsnIleTyAsnLys 1851
QY 8016 ATGAGCAAAAT--AAAAAT-----AAAGTAAATTAATCAACAGTATCAAAAC 8063
Db 1851 sLeuGlnAsnAspLysAsnLysTyGlyLysAspArgProLysTyLeuGlnLeu---- 1869
QY 8064 GTTTCGTTCTGCTGTGGGATGCTAATATAGAAATTTTGGAAAGCAATGAGCTGCA 8123
Db 1870 -----ArgSerAspTrpTrpAspAlaAsnArgLysHisIleThrAsnAlaMetThrCysAs 1888
QY 8124 AGCACAAGATGCAAAAATCTTTAGAAAAGCAAGATGAT-----GG 8168
Db 1888 nAlaProAspAspAlaLysPheLeuLysLysAsnProAsnAspThrSerGlySerSerSe 1908
QY 8169 ATTGGAACGCAATACATTAATCAAGATAACTGTGACATTAAGACGATCAACCTGTGGA 8228
Db 1908 rSerLysGlyIleMetThrThrHisSerAsnCysGlyTyArgAspLysGluProProAspTy 1928

QY 8229 TGATTATATACCTACAGCGTTTCGATGATGATGATGCTGATTAATTTTGTAAAC 8288
Db 1928 rAspTyIleProGlnProPheArgTrpMetGlnGluIlePheGluSerPheCysLysLe 1948
QY 8289 ACTGATGGAAGATTTGAAAATTTTAAATTAATCATGATCATGCT--AAAACATCGCA 8345
Db 1948 uLeuAsnGlnGluMetGluGlnPheGluLysThrCysGlyGlyLysLysLysAsnSerI1 1968
QY 8346 CAGATGCAAGATGATTAATGAAATTAAGTGTGACAGCTGTAAACGAGATGTCAGA 8405
Db 1968 eThrCysGlnAspAspArgAsnGlyThrAsnCysGlnAsnCysLysAsnGlnCysGluTy 1988
QY 8406 ATATATAAATTTGTTCTTAAATGAAATCTTATGATATACATCAATTAATTAACA 8465
Db 1988 sTyLysLysLeuIleHisAsnTrpLysLeuGlyPheAsp-----LysTyLys 2004
QY 8466 AGAATGATGATGACAAACCAATATAT-----ACAAATATCTCTATGATGATCA 8513
Db 2004 sGluIleTyAsnGlu--IleTyAsnAsnLysAspSerLysIleAsnSerAsnGluTy 2023
QY 8514 TGTTCAAAATTTTGTACAAAAGTTGAAACCTTTAAAGTGAAGT--TCTGTGAGAG 8570
Db 2023 rPheLysLysPheLeuGluLysLeu-----LysAspLysCysLysGluLeuAsnSe 2040
QY 8571 CTTTTCGAATATCTTCAATGAAACAGTACTGTTGATTAATTAATTAATTAATGA 8630
Db 2040 rSerAspLysCysLysAspGlnAlaThrHisCysThrLysTyLysPheSer---AsnSe 2059
QY 8631 TGTTCCTTCCAAATATACGAACATATGCTTTCGAGAAACACCAAAAGTTTAAAGAAC 8690
Db 2059 rGluAsnLysAsnHisAsnAsnTyAlaPheLysAsnProProGluTyGluLysAl 2079
QY 8691 TTGAGTTGTACACTACCTTCTTAAGAAATTCATGATTAATTTGCTTACCGATCA 8745
Db 2079 acLysLysCysAspAlaPro-----AspProLeuAspAsnCysProLysAspSerAlaTh 2097
QY 8746 -AACAAAGATGATTAAGAAATTAACAACTTTTACCTTCTGCTCGAAGATGATTA 8804
Db 2097 rTyGluLysAlaCysAsnThrLeuLeuProThrLysLeuGlySerLysThrPheAs 2117
QY 8805 TAATATATCTTAATTAATGAGACGATACCTTCTTAAATGATGATGATTAACAAAG 8864
Db 2117 nAsnAspAspAspSerTrpAspThrSerPheValGlnThrSerProAlaGAspAsnThG1 2137
QY 8865 TGTATGATTCCTCCAAAGAGACATTTATGACAAACCTTACCTGATCAATTAATTA 8924
Db 2137 yValLeuValProProArgArgArgGlnIleCysLeuLysAsnIleThrThr---LysLe 2156
QY 8925 TAGAAAAGGTATAA--GAAATTTTAAAAAAAACCTTCTTACTTCTGCTTCAGTCA 8981
Db 2156 uArgSerIleGluLysIleAspAspPheLysAlaGluLeuMetThrSerAlaTyAsnG1 2176
QY 8982 AGCAAAATGTTAGGTCAAAAATTAATTAATGGAAGAAAGTGTGCTTGAAGCAATGA 9041
Db 2176 uGlyLysLeuLeuGlyGluLeuTyLysLysAspArgAspAlaThrLeuGlnAlaMeLys 2196
QY 9042 ATATATGATGCAATTAATTCGATATTAATTAAGAACTGATATGATGACACTTCAT 9101
Db 2196 sTySerPheTyArgAspTyGlyAspIleValLysGlyThrAspLeuIleSerThrAlaPr 2216
QY 9102 ATCTGAAAATTAATTAATTAATTT-----GAAACATCAATTAAGAC 9143
Db 2216 oLeuAspLysLeuLysThrLysLeuAsnValLeuLeuLysGlyAspGlyThrAsnGluI1 2236
QY 9144 AACGCAAAATGCTAAACATGATGAGGAAATTAATAGACGTAGATGACAGCTATGT 9203
Db 2236 eLysGlnAspArgGlyLysTrpTrpThrGluAsnArgThrArgValTrpHisAlaMeLe 2256
QY 9204 ATGTGATATATAAATGCTACTTCAAAAAGTAAATTAATTAAGAGATGCTCAATTAAC 9263
Db 2256 ucysGlyTyTrpLysAlaAlaGlyLysIle-----GluGluLysArgAspCysSerLeuPr 2274
QY 9264 AAAGATGACAAACATATACGTTTCTGTTGTTTATTAATGAGCAAGCAAGCATG 9323

```

Db 2274 0-AspAspAsnThrHisGlnPheLeuArgTrpPheArgLysIleuSerGlnHisPheCys 2293
QY 9324 TAAGAAAGAAACATGTAAGTATCATTAATAAACAATAATCTCGTCAACAGAGA 9383
Db 2293 SAALysArgGlnLysLeuPheAsnGluValLysArgGluCys-----AlaSerAlaG1 2311
QY 9384 TAATTTGAACCGTCAGAAATTTAAGACAACTGATGTCAG-----AATGATAT 9434
Db 2311 nCysIleIleGluTyrGlyThrIleAspProValCysGluLysIleuSerGlnTyr 2331
QY 9435 TAGAAATATATATGCTGTAATATGATATAAATAATACATGGAATATATAA 9494
Db 2331 rArgAspTyrIleThrArgLysIleGln-----G1 2341
QY 9495 ATATAGCAATATCAAGATCTTCAGGTAAATATAGCAATTAACCATCTGAA--GA 9551
Db 2341 uTyrArgLeuLeuAsnTyrGlnTyrAsnThrAsnPheAsnGluLysLysAlaGluValTh 2361
QY 9552 AAATGTCACATCATATATAAATCAAAAGATCTCAATGCGCTTGGAGTTAATGATAT 9611
Db 2361 rLysAlaProGluTyrPheAsnAspLys-----CysAsnAspLysCysAsnCysIle 2378
QY 9612 AAATGAATATGTTACAGACAAATAATATGAATAATGATTCAAAGACTACTAA 9671
Db 2378 uSerLysTyrIle-----AspIleGluLysLysTyrLys 2389
QY 9672 AAAATATATATCTGGTTATATTTGTTGAAGTGAACACACAAATAATCATCTACTAGA 9731
Db 2389 sAsnMetTyrAspSerPhe-----AspAsnAsnAspLeuLysAsnLysCysIleCys 2406
QY 9732 TGAATAATTAAGAAGAGAGAGAAACAGTCTCTTAAGACCTCTAATTTCTTTACACC 9791
Db 2406 sArgGlnIleLysProLysArgPro-----ProLysLysValLysLysProGluLys 2423
QY 9792 CCATAGATATCTTTCTATCAAGACCTTTATCTCAACACATGAGTGAACAATATGA 9851
Db 2423 uHisThrProSerGluGlnAspThrProProleuProProLysProAspAspLeuPr 2443
QY 9852 TCCT-----AAAAATGATATATGAAGTAGTATCTGTTGTTAT 9893
Db 2443 oProProAlaGluGluProPheAsnArgAspIleLeuGluLysThrIleProPheGly 2463
QY 9894 TGTATCGCGCTAGATGTTGATAGCGCTTCATTTGATGAAGAAATAATCAATCGCTCT 9953
Db 2463 eAlaLeuAlaLeuGlySerIleAlaPheLeuLeuLysLysThrLysSerSerVa 2483
QY 9954 G---GACTTGTTCGCTAGTACGAATATCCGCAAGAGAGATGATGATGCTTACGTGGA 10010
Db 2483 lGlyAsnLeuPheGlnIleLeuHisIleProLysSerAspTyrAspIleProThrLys 2503
QY 10011 ATCCAAATAATAGTATCATCATATAGAAAGTGTTCATATTAAGCAAAACATATATATA 10070
Db 2503 uSerProAsnArgTyrIleProTyrThrSerGlyLysTyrArgGlyLysArgTyrIle 2523
QY 10071 TATGAGAGAGATATCT-----AGTGCAGATGAGATTAATATATGTCGACTTATTC 10121
Db 2523 rLeuGluGlnLysAspSerGlyThrAspSerGlyTyrThrAspHisTyr----- 2538
QY 10122 TTCCCTGATATATCT---TCATCCGAAAGTGAATGATGAAGATTTGATTAATGATAT 10178
Db 2539 ---SerAspIleThrSerSerSerGluSerGluTyrGluGluLeuAspIleAsnAsp 2557
QY 10179 ATATATACAGTAGTCTTAATAATATAAATGATATAAGTAGTACAGTACCAATCAAA 10238
Db 2557 eTyrValProGlySerProLysTyrLysThrLeuIleGluValValLeuGluProSerG 2577
QY 10239 AAGGATATACCAAGTAT-----GATACACCAAGT-----AA 10271
Db 2577 yAsnAsnThrThrAlaSerGlyLysAsnThrProSerAspThrGlnAsnAspIleGln 2597
QY 10272 TGATACACACGATGAGATTAATTTATGATGATGATGATGATGATGATGATTT 10331
Db 10331

```

```

Db 2597 nAspGlyIleProSerSerLysIleThrAspAsnGluTyrAsnThrLeuLysAspGluPh 2617
QY 10332 TGATATCAATATTTA-----CCAAATTCAGAACCAAAATTAATATACAAAGTGC 10382
Db 2617 eIleSerAsnMetLeuGlnAsnGluProAsnThrGluPro-----AsnMetLeuGly 2635
QY 10383 AGATATTCATGATACAGAACCTTAATCTTATATTTGATATCTCGAAGAAAACC 10442
Db 2635 rAsnValAspAsnAsnThrHisProThrThr---SerArgHisAsnValGluGluLysPr 2654
QY 10443 TTTTATTAATATCATCATGATAGGATTTATATCTGGGAAAGAAATAGTATATAT 10502
Db 2654 oPheIleMetSerIleHisAspArgAspLeuTyrSerGlyGluTyrSerTyrAsnVa 2674
QY 10503 TATATGATGACTAATACATTAATGATATTCATGAAATGCTAGAAATGATCTTATAG 10562
Db 2674 lAsnMet-----ValAsnAsnAspIleProIleSerAlaArgAsnGlyAsnTyrSe 2691
QY 10563 AGGTATAGATTTAATTAATGATGATCTAGTGTGCAAAACATGATGATATATGATGAA 10622
Db 2691 rGlyIleAspLeuIleAsnAspSerLeu---AsnSerAsnLysValAspIleTyrAspGlu 2710
QY 10623 GTATTG 10628
Db 2711 LeuLeu 2712
RESULT 6
T28161
Hypothetical protein FCR3-variant1-1 - malaria parasite (Plasmodium falciparum) (fragme
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28161
M: Hernandez-Rivas, R.; Mottel, D.; Sterkers, Y.; Peterson, D.S.; Wellens, T.E.; Scher
Mol. Cell. Biol. 17, 604-611, 1997
A:Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.
A:Reference number: Z20483; MUID:97154495; PMID:9001213
A:Accession: T28161
A:Status: preliminary; translated from GR/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2647 <HER>
A:Cross-references: EMBL:U67959; NID:91794235; PID:91809295; PIDN:AA047438.1
A:Experimental source: strain FCQ27/PNG
A:Genetics:
A:Introns: 2158/3
A>Note: FCR3-variant1-1
Alignment Scores:
Pred. No.: 9.18e-127 length: 2647
Score: 2727.00 Matches: 922
Percent Similarity: 34.82% Conservative: 399
Best Local Similarity: 24.30% Mismatches: 810
Query Match: 14.02% Indels: 1664
Gaps: 129
US-10-087-013-1 (1-10628) x T28161 (1-2647)
QY 91 AGTCACAAAGTGCAGAAATTTGTAACGTTATGCAAAATATATACATTCATCA 150
Db 9 ThrAsnLysSerAlaLysGluValLeuAspGluIleGlyLeuThrIleGln----- 25
QY 151 AAATATGCAAAACACATGTGGATTCG-----TTGAAGGGGATTTG 192
Db 26 -----LysLysAlaHisSerAspAlaAspThrPheArgSerGlnLeuLysGlyAsnPh 43
QY 193 ACGAAGACAGAAATTCGTGGTGGT-----CCTTCTACGCCAGTAATTAAG 237
Db 44 GlyGluAlaLysPheTyrAsnGlyGlyGluIleMetGlnProAsnSerLysLeu----- 61
QY 238 CATTAATTTATTAATCATATCATGTAATTTAGATCATAGCAATCACTAATTTACG 297
Db 62 -----CysGluLeuAspHisThrIleAspThrAsnValThr 73
QY 298 TATGATGATGTAATTTGAGACATCCTTCCATGTTAGAGAACAAACGATTTGATGAA 357

```



```

OY 2395 AGAAGAGCTTCAAGCAATTTATGATTAATGATTAACATCTGTAATCTT 2454
Db 795 ThrIleuLeuLysAsp---IleCysSerIleThrAspHisSerAsnAlaLysArg 813
OY 2455 GGTTCCTCAATGACCATGTGATGAGCAAGGCAACAGGTGATGATTAACAACAGATT 2514
Db 814 GlyIleThrSerGlnProCysLysArgLysAspSerLysValAsnValLysAsnArg--- 832
OY 2515 GTCTAGCAATGCAATGGAGATGATCCGAAACATGCGTAAAGATTCAGCAAGATGT 2574
Db 833 -----ArgTrpMetAspThrAlaGlyPheIleSerAsnThrTyrLysAspIle 848
OY 2575 ATTATGCTCTTCAAGACACATATATGATACATCAATTTGAAACATTTCAACAGCAT 2634
Db 849 TyrMetProProArgArgGlnHisPheCysThrSerAsnLeuGluTyrLeuGlnThrThr 868
OY 2635 GATCACCACCTTAATGATTAATTTGAT-----GATTATGTTAATTAATCTCTT 2685
Db 869 AsnLysLeuLeuAsnGlnLysAsnAspIleAsnGlnAsnProAsnIleIleAsnAspSerPhe 888
OY 2686 TTGGGGATGTTCTTCTATCAGCAAAATATGAGCAAAACAGATTAATACGATGTTAA 2745
Db 889 LeuGlyAspValLeuPheAlaIleAsnTyrClnAlaAspPheIleLysLysMetTyrAsn 908
OY 2746 GAAAAAGATTAACCTAAAGGCCCCCAAGAACTACTGACCCAAACACACAGACACTATC 2805
Db 909 LysGlnAsn-----AspTyrLysAspAsnAlaThrIle 919
OY 2806 TTCTGAGCTTATACGATTATTTGAGATATAGTGTATATATTCAGAGAGA----- 2859
Db 920 CysArgAlaMetLysTyrSerPheAlaAspLeuGlyAspIleIleGlnArgGlnHisIle 939
OY 2860 -----GATCTGTGGAAAGAAACGGTGCATG 2886
Db 940 CysArgIleMetIleValGluArgValLysHisGlnIleSerGlnArgAsnPheLeuIle 959
OY 2887 GTAACCTCAAGACATTTGGAACTGTTTGTGTAATATACATTAAGTCATCAAGGC 2946
Db 960 LeuSerLysLysAsnIleLeu-----AlaPheLysGlnIle----- 971
OY 2947 AAAGCAATGATTAATATATGATGATGCCCCCAATATTTAAATTTAGGAAAAATTGG 3006
Db 972 -----TyrLysGlnAspThrPro---TyrThrLysLeuArgGlnAspTrp 985
OY 3007 TGGGAAGCTAATAGAGCCAAATATGAGGAGCAATGATGATTAATTAATTTGAAG 3066
Db 986 TrpGlnAlaAsnArgLysLysIleTrpGlnAlaMetGlnCys----- 999
OY 3067 GATTAATCGGACACCAATCAACACAAAGTATGTTGGGATATAGTCATACACA 3126
Db 1000 -----ProThrProAsnGlnLysSerPheProCysLysSerTyrHisIleGly 1014
OY 3127 TTGATGATTAATATCCACAAAAAATTAAGATGATGACCGAATGGGCAATGTACTGC 3186
Db 1015 LeuAspAspTyrIleProGlnArgLeuArgTyrTrpMetThrGluTrpAlaGluTrpPheCys 1034
OY 3187 AAGTCGCAAAAAAGAGATGATTAAGTTGAAAGAGAGTGTAAAGAGTGTAAAGATTAAG 3246
Db 1035 LysGlnGlnLysLysGlnTyrClnLysLeuValSerAlaSerAsnGlnCysLysAsp--- 1053
OY 3247 GATTAATGCTCAAGCTGTACAGAAAGAGAGTGTACAGTGTACAGAAAGTGTACAG--- 3301
Db 1054 GluArg-ValLysValAlaArgIleArgValHisAsnValGlnArgAlaCysLysHisVal 1073
OY 3302 -AAGCTGTATGATTAATATGATTAATAGATTAATGAAAGAAACAATGGAATTAATA 3360
Db 1073 LysIleIleLysAsnLeuLeu---IleHisGly-----LysGlnGlnTrpAspLysMet 1090
OY 3361 TCAGATAATCAAAAGATTAATCAATGACACACAAATGCTGTAGTAATAGGATTT 3420
Db 1091 GluIleLysTyrLysLeuLeuTyrLeuGlnAlaGlnThrThrAlaAlaAsnGlyGlyPro 1110
OY 3421 GAGCTTCAGTACTGCCAAAAATCATATAGACAGAAATGTAATGTAATTTTTCGAA 3480
Db 1111 AspThrTyrSerIleLysLeuValAspGluAsnGluLysProValValAlaAsnPheLeuPheGlu 1130
OY 3481 TTATACCAACAAAAATGTTGGCAAA-----AGTAAATAAA 3513
Db 1131 LeuTyrLysGlnAsnGlnLysIleGlnLysAsnProArgAspThrProArgAlaLysArg 1150
OY 3514 AGTGTACTAGTGTAGTAAGAGTGTGATGCTGATGCTACTAACCCAGTAAAGATTTGGA 3573
Db 1151 SerLysArgGluThrAlaProAlaSerValAlaLysAsnAspValLysSerThrAlaAla 1170
OY 3574 GCATATCTGCAT---GATCAGCAAAATTTGATGATGTGCATCAACAAAGAGTTTGT 3630
Db 1171 GlyTyrValHisGlnGlnIleMetGlyProHisMetGlnCysLysThrGlnThrGluPheCys 1190
OY 3631 GATGAAAAAAGTATGATGATTAAGCAAAATATGCTTAAAGATTAACCAACAGAGAC 3690
Db 1191 ---GluLysThrAspGlnGlnIleTyrAsnGlnLysAsnTyrThrPheLysAsnProProGln 1209
OY 3691 CATGATGTCGCTGTGTTGTAAGAGTGAATGCAACGCAAGGTAACAGTAAATAACG 3750
Db 1210 TyrLysAspAlaCysIleCysAsnThrArgProProLysGluAspSerArgLysArg 1229
OY 3751 AAAAAA-----AAAGCGAAGAAAGATTCG 3777
Db 1230 SerGluAspSerAspGlnGlnLysValLysGlnThrLysValGlnLysAlaThr 1249
OY 3778 GAA----- 3780
Db 1250 GluAspAlaValAspThrGlyProProProAlaProLysGlnAlaThrThrThrLeuAsp 1269
OY 3781 ---TGTAACAGTAAATGATTAATTAAGAAAAACGTGAAAGAAACAAGTAAAGAT 3837
Db 1270 ValCysProIleValAlaGlyValLeuThr-----LysGlnAsnLeuGlnAsn 1285
OY 3838 TGTCAATCCAAAAAAGAAATAGATGATATCCC---GATTCGCAATGC----- 3882
Db 1286 AlaCysProThrLysTyrGlyProLysAlaProThrSerTrpLysIleProThrGlu 1305
OY 3883 -----GGAATATTAATTAAGTGAAGAC 3906
Db 1306 LysThrAsnAlaIleThrGlySerGlnGlySerSerGlnLysAlaLeuGlnArgAla 1325
OY 3907 CCTCGT-----GTCGTGATG 3921
Db 1326 LysArgAlaThrValGlnSerGlySerProValThrSerAsnSerGlySerIleCysIle 1345
OY 3922 CCCCCTAGAACAAAAAGTTATGCGTA-----CATTTCTTGCAAAATGATTAATGAA 3972
Db 1346 ProProArgArgArgArgLeuTyrIleGlnLysLeuHisAspTrpAlaSerGlyAsnThr 1365
OY 3973 ATAAAAAATTAACAATCAAAAGT----- 3996
Db 1366 ValValSerGlnAlaGlnThrProGlnGlyLysThrSerSerProSerGlyLysGlu 1385
OY 3997 -----AATTTAAAGAGCTTTCATCAAAATCTCAGACAGCAAAATCTTCT 4044
Db 1386 ThrProSerAspLysLeuArgThrAlaPheIleGlnSerAlaIleGlnThrPhePhe 1405
OY 4045 TCATGATTAATTAATAAATTAAGATGTCGAAGAGAAATCACTCGATTAAGAATTAATA 4104
Db 1406 LeuTrpAspArgTyrLys-----LysGlyLysAlaIleAlaLysLysGlnLys 1421
OY 4105 GAAGGCAAA----- 4113
Db 1422 LysLysGlnMetValAspTyrSerProLeuSerThrAlaAspProHisAsnAsnProVal 1441
OY 4114 -----ATTCTCCCGCATTT 4128
Db 1442 SerLeuValIleAlaProAsnProAsnTyrAsnLysThrCysValIleProProProPhe 1461
OY 4129 TTGAGATCCATGTTTCAACATTTGGAGATTAAGAGATTTTAAATTGGAACAGAT--- 4185

```

||||| ||||||| ||||||| ||| ||||| |||||
Db 1462 LeuArgGlnMetPheTyrThrLeuGlyAspTyrAlaAspIlePhePheGlyLysAsp 1481
QY 4186 -----ATATCAAAAGGTCATGCTGAG-----GGAAGTAACTAAAA 4221
Db 1482 IleValIleAspThrLysAsnGlyAspIleAlaGluArgLysLysIleLys 1501
QY 4222 GACCAATAGATCTCTTTTCAAAAATGTCACCAAAAATCTCTTAATGGAAAAACACGC 4281
Db 1502 AspAlaIleGluArgValLeuLysAsnAlaAspSerGlnProPheSerAspGlyLysArg 1521
QY 4282 CAGCAATGGGAGCAGACATAGTATGAGATATGGAACTATGGTATGTCAGTATGTA 4341
Db 1522 GATThTTPTrpGluGlnAsnGlyLysIleThrPheAsnGlyMetIleCysAlaLeuThr 1541
QY 4342 AAAATGGGGCAAAAAAAGATGATTTTACCGAAAGCTACGGTATCAACAACATCAATTT 4401
Db 1541 -----TyrLys----- 1541
QY 4402 AGTACAAAAAGCACACTTTGGAGAAATTTGCCAAACGACCCGATTTTACATGGCTA 4461
Db 1541 ----- 1541
QY 4462 ACCGAATGGTACGACGACTATGCTATACGACAAAAATATTTGAAGATGTGACGAA 4521
Db 1541 ----- 1541
QY 4522 AAATGTAACTCAATGACCAATTTGAAGTGTACAGAAATGTAAAGAAATGCGAGAC 4581
Db 1541 ----- 1541
QY 4582 TACGTTAAATATATGAAAAAAGAGTGCATCCACAGATAAATTTACAGAGAT 4641
Db 1542 -----TyrLys----- 1543
QY 4642 GAACGGACAAAAAAGATTCGATAGACACACATGTTGTTTACAGACTATCT 4701
Db 1543 ----- 1543
QY 4702 GGAACGAATGCAACAGATTAATTGAACAGAAATTTACTGCTAGTGTGTATTAACCT 4761
Db 1543 ----- 1543
QY 4762 GGAAGTCCCTCTGTGTACAAAATAATATCAATTTGTAGAAAAACAGCTTACTATGAT 4821
Db 1543 ----- 1543
QY 4822 GCCGCAAAACATTTGGGTGCACAAAATTTATGAAAATGACGCAAAATATCAAT 4881
Db 1544 -----GluLysAspGluLysGlyThr----- 1550
QY 4882 TCGAGTAAGATTAAGTCAAAAGATTAGTAAGAGAGCAACACAGGTCTATTAAAGTGG 4941
Db 1550 ----- 1550
QY 4942 CAAAACAAAGGCTTAATCACTCAATTAAGTAAGATGACAGATGTGCTTTT 5001
Db 1550 ----- 1550
QY 5002 CCTTCTGCTGACTATATATGTTTTCATGCATGTGATGGCAATTTATACAGATCCGAA 5061
Db 1551 -----Pro 1551
QY 5062 GTTAAAGATGAAAAATGGTTGCAAAAAAGATTGATGAAATGGCGCAACGAAAGGTAC 5121
Db 1552 LeuLysGlnAsnGluLysLeuLysSerAlaLeuTrpAsp----- 1564
QY 5122 AATTGGGTCAATCTACAAAGAAAAAAGAAAAAGAAAAAGAAAAATTAACGTCGATGCG 5181
Db 1565 -----GluLysAsnLysLysProLysAspGln----- 1573
QY 5182 CACAAATATCTTATGAGGTCCCGCTTGTAGTGTATGAATATAGTTTATGATTTA 5241
||||| |||

Db 1574 -----LysTyrGlnTyr----- 1577
QY 5242 AGAGATATATATCTAGTATGATTAATTTGGAAGATGAAAAACAAAAGACGAGAAAT 5301
Db 1578 -----AspLysValLysLeuAspGluAsn 1585
QY 5302 TTGACAAAAATATTAAACAAAAATGACACTCA-----GTGGCAAAAGAGAT 5349
Db 1586 -----SerGlyThrSerProLysIleValValProAlaProLys 1598
QY 5350 GATAGTACTACAGAAATCCCGTAGTACTGCCGAAATTTTTCGAAACGAAATAAG 5409
Db 1599 ProThrThrThrPheProProProProSerProThrSerPhe----- 1612
QY 5410 GAATGTGTGTGAACGCAATGATATGCGGTACAAACGTGTGAGGATGATGAAATAGT 5469
Db 1612 ----- 1612
QY 5470 GGAATATGTCAAGAAAGTGAAGATCTAAAAAATGTGTTCTGTACTTACATGAT 5529
Db 1613 -----SerArgPro----- 1615
QY 5530 GATTATCCTATGGGAAAAATTCGCGATGAAGTACTGCGTATCAGTTCTTGATGGTT 5589
Db 1616 -----ProTyrPheArgTyrLeu 1621
QY 5590 GCCGAATGGGTGACAGATTTTTCACAAATTAAGAAAAAGAAATTTGGAATTTGTAGCG 5649
Db 1622 GluGluTrpAlaGluThrPheCysArgGluArgLysArgLeuGluLysIle 1639
QY 5650 GCGTGTATGATTAATCTGTGTGATTAAGAAATTAAGAAATTAAGAAATGATCA----- 5703
Db 1640 -----LysValGluCysMetAspGluAspGlyLysLysGlnLysCysSerGlyAsp 1656
QY 5704 -----GATGCGTGTACCAATTAATAAATTTATATAGAGAGGAACACAGTATGAAAAA 5760
Db 1657 GlyGluAspCysGluGluIleArgLys----- 1665
QY 5761 CAAATCAAAAAATATGTGTGAATTAAGACAAATAATATTCGACATCTGTGCAAAA 5820
Db 1666 -----GlnAspTyrSerThrValArgAsp-----PheTyrCys-----ProGluCysGlyLysT 1682
QY 5821 GATCAGAGAGCGCTGCGCAATTTTGTACAAACATTTAAAAAATTTGTGAATTA 5880
Db 1682 TyrCys-----ArgPheTyrLysArgTyrIleGlyLys----- 1692
QY 5881 AGTGAAGATTGGAATATAGTATAGTAAGATGTGTCCACACAGCGATTAAGTGT 5940
Db 1692 ----- 1692
QY 5941 AATAGTCAAAATATGCCCGCATATTAGACATGAACCAAAAGAAATTTGAAGAAAGTGT 6000
Db 1692 ----- 1692
QY 6001 AATTGTCAAGTCCACAGAGTCCACACGTTGTACGAAGGAAACCCGTACCAAGGCTA 6060
Db 1692 ----- 1692
QY 6061 TCACTGATATCAAAAGCGACGCGATCGAAAAAAGAAAGCAAGCGCGCTACAAAA 6120
Db 1693 -----LysLysAspGluTyrAspLysGlnLysGlu-----AlaTyrAsn 1706
QY 6121 CAGCCGAAAAAAGTGAATAATCTAACACAAAGAAATGGAGACACAAACAGCAACCCGACGA 6180
Db 1706 snGlnLysThrAspAlaArgArgAsn-----AsnAspAsnAlaPheS 1721
QY 6181 GCAGCACAACAAACAGAAACAGACATCAACGACAAACAAACAGATGTGACGTGGCG 6240
Db 1721 erThrThrLeuAspTrpThrCysThr-ThrAla-----Gly 1731
QY 6241 ACAATGTAAAGGCCATTTCTTGAATTAACAGATAGCAGGGGTGAATAGAGGGTGT 6300
Db 1732 AspPheLeuGlnThrLeu----- 1737

```
QY 6301 AATCCAAAACGATGACATATCTAAATGGGGTGTATGTAGGTAAAGTAAAGAA 6360
Db 1737 ----- 1737
QY 6361 AATGAAATGGCATATGTATGCTCTAGAGAGAAAAATTAATATATATACAA 6420
Db 1738 ---LysasnGlyProCys----- 1742
QY 6421 TATTTAAATTAAGAACTGAAATTAAGCGTACATATATTAAGAGCGTTTATTTAA 6480
Db 1743 -----LysasnAspAsn----- 1746
QY 6481 TGTGCACCAATAGAACTCAATTTTGTGTAAATATATTAATGAAATCCTGCAGCA 6540
Db 1746 ----- 1746
QY 6541 GAAATGATTCGAAATGAGACAAATTCAGATGATTTAAAGAAATATGTATATACA 6600
Db 1746 ----- 1746
QY 6601 TATGTGATTTATAAGATATGTTTGTGAACTGATATTTCTATGATATAAAATTAATA 6660
Db 1747 -----ValAspAspSerCysGlyIleAsn----- 1754
QY 6661 ACTGTAAACAATAGTGTACACACATTCATGAAATATATAGAAAAACAGATAAA 6720
Db 1754 ----- 1754
QY 6721 AAAAAAGATGAAAGATTTAGTAAATATTTTGGAGAAAAATTAATTTATTTGGAA 6780
Db 1755 -----LysIlePhe----- 1757
QY 6781 GGAATGATATATGATTAATTAATATCTATCTACAGACGAAAGAAAGAAAAATTAGA 6840
Db 1758 -----AspGluAsn-----Gly 1761
QY 6841 GATTAATACAGTACATGACATGACAACTGACGCTTCCTGAGAGTTTGTAAA 6900
Db 1762 ---AspThrPheLysIleThr----- 1767
QY 6901 AGGCCCAATTTTGTGATGTTTCACAGAAATGGCAGAAATTTGTATTAAGAGAG 6960
Db 1768 -----GlnIleCysGly----- 1771
QY 6961 GAACAGTGTAAATTTGAGGCGGCTGTAGAAATATGAGTATATGCTATGATGAC 7020
Db 1772 -----ThrCysSerIleuAsnGlyPheLysCysAsnGlyAspAsp----- 1784
QY 7021 GGTAAAGACAAAGATGTGACAGGCGTGTATACATATCAAAATTTTATTAAGAGTGG 7080
Db 1785 -----CysArgValAlaGlyThrAsn----- 1790
QY 7081 AAACTGATATGAAAGACAAAGAGAAAAGTTCAAAAAGATTAAGTGGCAAAAAGTAT 7140
Db 1790 ----- 1790
QY 7141 AAGGATATCTCTACTGAAAGACATATAGAGAGCAACATGTCATGATATTTA 7200
Db 1791 -----ValIleCys----- 1793
QY 7201 AACTGAAATTAAGAAATTAATGTGCAATTAAGATGTCTGTATGCAAAAACCTTCT 7260
Db 1793 ----- 1793
QY 7261 TCACAACTACAAAACAAACAAATCACAATCATCCGATGCTATGATATGCCAGAA 7320
Db 1794 -----AsnGlySerAsnArgThrThrIleThrAlaAspAspIle----- 1807
QY 7321 TCGGTGATATGTTCTCGAAGAAATTTAAACAAGTGTAGTGTCTGAACCTTCAAAAAG 7380
Db 1808 -----LysAsn 1809

QY 7381 GGATCTATGATTCATCAAAAAAATTTACTGAACTTAAATACCTTATGATTTGTAGAG 7440
Db 1810 Gly----- 1810
QY 7441 AAAGCAGATATTTATTTATCTAAAGACAGACAAAATATATG-----GATATT 7488
Db 1811 -----GlyAsnSerAlaGluIleAsnMetLeuValSerAspAspIle 1824
QY 7489 ACCTTAAGAAAAATTTTATACCTATGAGTCTACAAAGAAAAAGAAAGTAAAAATAGT 7548
Db 1825 AsnSerGlyAsnGlyPheAsnAspLeuGluAla----- 1835
QY 7549 TGGACTAATATATATCTTGCATCTCTAAGAAACCTTATGACCTGATTAATATAGTA 7608
Db 1835 ----- 1835
QY 7609 AGAAGAAACCTTGTGAAATATAGAGAGAAAAATCCTTTAAG-----GTAGATTATGAA 7662
Db 1836 -----CysIleAsn-----AlaAsnIlePheLysGlyIleLysGluAsnLys 1849
QY 7663 TGGAAATGTTACAAAAATTTCAAGTTCATCTACGAGAAAAAGAGTATGTACCTCCA 7722
Db 1850 TrpLysCys-----ValIlePheCysLysSerAspValCys----- 1861
QY 7723 AGAAGAAACATATGTGCTTAAGGAATTTAGATGAATTAATTAATGAAAGCTTAAGAT 7782
Db 1861 ----- 1861
QY 7783 AGTATTTATCTCTTAAATATGTTGCTGTCGACGACGAAATGAGAAATAGACATATA 7842
Db 1861 ----- 1861
QY 7843 AAAAACTTCAACTCAGAGAAAGGGTGCACATGATCCAAATATGATACTGAAATAT 7902
Db 1861 ----- 1861
QY 7903 AGTTGCTGATCTGGGTGACATAGTTAGAGAAACAGATATGTTACGAATTTGTGTAC 7962
Db 1861 ----- 1861
QY 7963 TTACTCCGTAGAAATTAATTAATTAAGTTTTGAATCATATATGAAAAATGAGA 8022
Db 1861 ----- 1861
QY 8023 AATAAAAAATTAAGGTAGAAATTAATACACAGCATGTACAAAAGTTTGTCTGTGTGG 8082
Db 1862 -----GlyLeuLysLysAsnAsnAspIle----- 1869
QY 8083 GATGCTATAGAAAAAGATATTTGAAAGCAATGACGTGCAAAAGCAGACAGAAATGCAAA 8142
Db 1870 AspGluAsnGlnIleIleLeuIleArgAlaIleu----- 1880
QY 8143 CTTTGTGAAGAAAGAAATGATGATGATTTGAACGCATTAACATTAATACAAAGTAACTGT 8202
Db 1881 -----PheLysArg----- 1883
QY 8203 GCACATTAAGACGATCCACCTGTTGATGATTAATATACCTCAAGGTTTCGATGATGACT 8262
Db 1883 ----- 1883
QY 8263 GAATGCTGAAATTAATTTTAAAGCAGCTGATGAGAAATTTGAAAAATTAATAATCA 8322
Db 1884 ---TrpLeuGlnIleArgPhe-----LeuAspAspIleArgAsnLysIleArgLysLys 1898
QY 8323 TGTGATCAGTGTAAACATCTGACAGATGCAAGAAATGATATGATGAATAATAGTGTGAA 8382
Db 1899 LeuAsnProCys-----IleAsnAsnGlyGluLysAla 1909
QY 8383 CAGGTGTAACGAGATGTCAAGAAATTAATAAATTTGTCTTAATGAAGAAATCTATATTC 8442
Db 1910 IleCysThrAsnGlyCysValGlu-----GlnIlePhe----- 1919
QY 8443 GATATACAAATCAAAATTAATCAAGAAATTTGTATGAAACAACAAATATATCAAAAATCTCT 8502
```


[illegible]

D	2017	ThSerArgL---	lleProGlnGlnProGlySerGluValSer--	ProAsnHisIleGluLysP	2034
Q	9577	AAAGATTTCTCAATGCCGCTTTGGAGTTAAATGATATA-----	-----AATGAA	9618	
D	2035	GIUGIUGInProLeuGluGluGluLysThrValGluHisProGlyS	IleGlySerAsp	2054	
Q	9619	ATAGCTACAGGACCAAAAATTAATGAAATATGAAATTCAAAGAGTCTATAAAAAATTA	9678		
D	2055	ValLeu-----	LysHisIAsnHisAsnGlnArgAsnGlnGluArgLeuValLys-As	2071	
Q	9679	TATCTGGTTTATATTTTGGTGAACATATACACACACAAAAAATCATGCTAGATGGAAT	9738		
D	2071	nProLeuValGlnProThrLeuLysArgLysLysLysLysLysValArgLysLysI	2091		
Q	9739	ATAAAAGAAGAACCAAAACAGTGTGCTCAAGACACGTATCTTATTTTACACCCCATGTA	9798		
D	2091	e---LysLysLysAsnGlnAspPheHisProArgHis-Leu-----	ProGlySg	2106	
Q	9799	GATTTCTTCTATCAAGCACCTTATTTCTACACACATCGAGTGGCAATATGATCT---	9855		
D	2106	LysAlaPheIleAsnThr-----	AsnThrProLysThr-LysThrProProSer	2121	
Q	9856	-----AAAAATGAT-----	-----AATATGAAAAGAGACATGCTGCTGTGTATTT	9894	
D	2122	SerGlyLysAsnProTyrPheLinhIsProAlaValIleProAlaLeuValThrSerThrLeu	2141		
Q	9895	GTATCGCGGTAGAGTTTG-----	-----AATAGCGCTTATTCATGACGAAGAAAATTC	9942	
D	2142	AlaTyrSerValGlyIleGlyPheAlaIaIaPheThrTrpThrPheLysLysLysThr	2161		
Q	9943	AAATGCTGTGAGCACTTGTGGCGTACTGAAATATCCCGCAGAGAGATGGAATGCTT	10002		
D	2162	LysSerThrIleAspLeuLeuLeuSerLeu---IleProLysSerAspLysIlePro	2180		
Q	10003	ACGTTGAATCCAAAATAGACATACATACATAGACATGAGCGTCCATATAGAGCAAAAACA	10062		
D	2181	ThrLysLeuSerProAsnThrGlyIleProGlyTrpThrSerGlyLysArgLysArg	2200		
Q	10063	TATATATATATATGAGAGAGATACT-----	-----AGTGAAGATGAATAAATATATGTGG	10113	
D	2201	TyrIleLysLysGluGluLysPserLysThrAspSerGlyTyrThrAspHisLysr-----	2218		
Q	10114	GACTATCTTCTCTGATATTACT---	TCATCCGAAGAGTGAATGAAGATTTGATATT	10170	
D	2219	-----SerAspIleThrSerSerSerGluSerGluTyrGluGluMetAspIle	2234		
Q	10171	AATGATATATATGTCCACGATGCTGCTCAATATATAAACAATTCATGACGAGTACTAGAA	10230		
D	2235	AsnAspIleLysValProGlySerProLysTyrLysThrLeuIleGluValValLeuGlu	2254		
Q	10231	CCATCA-----	-----	10236	
D	2255	ProSerGlyLysLeuSerGlyAsnThrIleProThrSerGlyAsnAsnThrHisProThrMetSerArgHis	2274		
Q	10237	-----AAAAGGATATATCCACAGTGAATGATACACCAAGTAAATGATACACACGTCAGAA	10290		
D	2275	AspThrGlnAsnAspIleProThrSerAspThrProProPheIleThr-----	2290		
Q	10291	AGATTATATGATGATGATGAATGAATGAACATGATATTTGTATCTTCATATTTTACCA	10350		
D	2291	-----AspAspGluThrPasnThrLeuLysHisAspPheIleSerAsnMetLeuGln	2307		
Q	10351	AATACAGAAACCAAP-----	-----AATATATTAACAAGTGCAGATATTCACATGATACAGAA	10404	
D	2308	AsnGlnProLysAspValProAsnAspLysThrThrSerGlyAsnSerSerThrAsnThrAsn	2327		
Q	10405	CCATAACTTATATATTCGAT-----	-----	10425	
D	2328	IleThrThrThrSerArgAspAsnValAspAsnAsnThrHisProThrMetSerArgHis	2347		
Q	10426	AATCGTGAAGAAAACCTTTTATTTATCTATCTATTCATGATAGAGGATTTATATCTGGAAA	10485		
D	2348	AsnValAspGlnLysProPheIleThrSerIleHisAspArgAsnLeuLysThrGluGly	2367		

QY	1528	AATGAACATATAAACCTCAGTCGGTGAAGCCCTACTAAATATACAGCTCCTTTATAGT	1587
Db	511	GLYLSARGLYRSNLLIEPOLYSGLYNTRLYSHLSASNAVALLEPROVALLEUSERPHE	530
QY	1588	GGTAATGAACAGGTGATATTATACACAAAATTAGAAAATTTTGTACACGCTCAACTAAT	1647
Db	531	GLYASPGILNHLYSGLNLEILLIEGLILULYSLEGLUGLNPHCEYALAGLUSERASNSER	550
QY	1648	TACAAAGATATAAATATATCAAAAATGCGAATCCTATTATTAAGATCAAAATATTAATATGA	1707
Db	551	ASPSESRILYSLNETHRGUGLGNTRPLYSCTSYTYRGLYASPLYSGLUTRYGLUVAL	570
QY	1708	TGTAACTGGAACAAAATATCTGAAATCAATTAATGATTAATCCTAATAATATATATTT	1764
Db	571	CYSNTRLEUGLUNASNAAGSNLYSSERGLUGLUNASPROGLUGNLEGLNLYSTRPHE	590
QY	1765	CATATATTTTTCATATTAATGCGTTACATATTTATTAAGGATATATTAAGTCGAATGAC	1824
Db	591	HISASNPHEPHELYRPHEPTRLIAAGNLSLEULENASNSERILEGLUTRPARASP	610
QY	1825	AAACTTAAACTGTATTAATATATATACAAACACG---CATGTATGATGATGATGATACGA	1881
Db	611	LYSILEASNSNDYSILIEGLULYSALALYSGLUGLYLWSCYLSYASNGIDCYSLYSTR	630
QY	1882	AATTCCTTATGTTTTCACAGATGGTTTAAACAAAAAGAGAAGATGATATGATTAAG	1941
Db	631	ASPYSGLYSPHEGILNAGTRPILIEGLYLYSLYSGLUGLUTRPGLYGLNILEYS	650
QY	1942	AAACTCTTCACAAAAAAGATATACGACCAATGCTATTATAGTAAATATTAATATCTT	2001
Db	651	LYSHISPHLYS-----THGNASPGILYSESRILEPHECGLYASN--	664
QY	2002	TTTCAAGCTTATTTTAAATGATTAAGTATGAAATCTTGACAAAGATGAAGCAAAATGANA	2061
Db	665	---ASNTRYASPRHEVALLEUGLUNASNAVALLEASNILLEASP-----	677
QY	2062	GAACTTATGGAATATATAAAGAAAAAATGACTTTCCAAATTGGAAATATATAGG	2121
Db	678	GLIUNEPHEGLNASPILETRHGLUALATRYGLYASNSERGLNLYSLEGLNGLYILEYS	697
QY	2122	GACTATTTA-----GAGAAATCA	2139
Db	698	ASPHRILEUNALALYSLSYSTHTRGILNALALALASPRPALATHRGUGLNLYSASNTR	717
QY	2140	ATGAACTCTTTGTAGATCACTTAAAAAAGAACTCCACAGATATGTAAAGACATATATCA	2199
Db	718	ILASPRLEUNEPHEGLUTRYASPSERGLUGLUALAGLULWYCYGLYS-----LYSILE	735
QY	2200	AACGACAGATGTGAACATCCAT---AATGCAACAACAACCGGTGTAAACCTGCT	2256
Db	736	GLNGLUGLUCYSGINPOLYSLYSPROTHLYSVALARGASNPCCSYTYR-----	752
QY	2257	GGAGGACGACCAACCACTATAAATATATAAAGAATATAGCAATCTTAAAAAGACTGCA	2316
Db	753	GLYSNASNTHTRYASPRALALALAGLYLYSVALALAGLNLYSLEUGLNLGNLUALA	772
QY	2317	TACAGCAACCA---CGAATCTCGCTCTTCATTAATTAAAGCAAGACACAGCAAGCT	2373
Db	773	LYSGILGILNLEUNASPRAGSNASPSERATRGSERALALALEULYSALIASNALSERCTNLIY	792
QY	2374	ATATATTAACGTGGGGGTAGAGAGAAGAACTCAAGGACAAATTTATGTAGATATATGATA	2433
Db	793	LYSTRYSERASNGILNLYASPRTPASPRAPHELYSLYSASLEUCYSGLYLETRHCLN	812
QY	2434	AAACATTTCTAATGCTAATCTTGTTTTTCAATGACCACTGTCGATGCGAAAGCCACAGCT	2493
Db	813	LYSHISSEIRASNALA---ILIEGLYASPSERYASNPCCSYASNASNLYSGLY-----	829
QY	2494	GATGCTTATTAACAAACAAATTTGCTGATGAACTGAATGGAGAGTGATGCCGCAACATG	2553
Db	830	---LYSGLUGLWGPHEASNAVALIGLYGLULYSTPRLYSASNGLYGLUN---VAL	845
QY	2554	CGTAAAGATCAAGAGATGTTATATATGCTCTAGAAAGACGACATATATGTACATCAAT	2613

Db	846	LysMetSerHisThrAspLeuTyrLeuProProGluArgGlnHisPheCysThrSerAsn	865
QY	2614	TTGGACAAATTATACAAAGCGATGATCAACCACTTAATAGTAATATGTTGATCATTTACTT	2673
Db	866	LeuGlnHisLeuAsnThrLys-----SerThrGlyLeuThrSerAspLysAla	881
QY	2674	AATATATCCCTTTTGGGGGATGTTCTTCATACGCAAAATATGAAGCAACAAAGATATA	2733
Db	882	HisSerLeuLeuGlnLysAspValLeuLeuAlaAlaLys-----	894
QY	2734	CGAATGTATATAAGAAAGAAATTAACSTAAAGGGCCCCAAAGAAATACAGCAACAAACAC	2793
Db	895	-----LysGlnGlyLysAspLys-----ThrLysLeuThrGlnLysAsnAspAsn	909
QY	2794	CAGACAACTATCTGTCGACSTATAACCTTACAGTTTTGACAGATATAGCGATATAATTCGA	2853
Db	910	ArgSerSerHisCysArgThrMetLysTyrSerPheAlaAspIleGlyAspIleLeuArg	929
QY	2854	GGAAGAGATCTCTGGGAAAGAAAGCGGAGCATGTTAAACCTCGAAGAAATATGGAAACT	2913
Db	930	GlyThrAspLeuThrAspIleAsnGlnLysAspValThrGlyValGlnAsnAsnLeuLysAsp	949
QY	2914	GTTTTGGTATATACATTAAGTCACTCAAGGCAAGCAAGAAATGATTAATATATGAT---	2970
Db	950	IlePheSerLysIleThrGlnGlnLeuLysLysGlnHisProAspLysPheAsnAspAsn	969
QY	2971	-----GATGCCCCCAATATTTTAAATTTAGGCAAAATTTGGTGGCAACTAT	3018
Db	970	AspLysTyrThrAsnAspSerLysHisThrLysLeuAspSerAspIrrPrrPrrLysAlaAsn	989
QY	3019	AGAGCCAAAGTATGGGAGGCCATGAATATGATTAATATTTAAGAGTAAATCGGCA	3078
Db	990	ArgAspGlnValTrrPrrLysAlaMetTrpCysProThrLys-----	1002
QY	3079	CACCAATCACCAAAAGTACTTATTGCGGATATAGTACATACACCATGATCATAT	3138
Db	1003	-----AsnGlnAsnIleGlnCysGly-----AlaThrProHisAspAspTyr	1016
QY	3139	ATCCCAAAATTAATATAGATGATGACCGAATGGGCAAGATGGTATACGCAAGTCAGAA	3198
Db	1017	IleProGlnArgLeuArgTrrMetValGlnTrrPrrAlaGlnTrrPheCysLysGlnHiser	1036
QY	3199	AAGAGATATGATTAAGTTGAAGAGAAAGCTTAACGAGTGAAGATTAAGATTAATGCTCA	3258
Db	1037	ArgLeuTyrGlnGlnLeuLeuAspAspCysGlySerCys-----ThrThrGlyLys	1053
QY	3259	GGCGTGTACGAAAGAGAGTGGTACAGGTGTACAGATGACAGCAAGCTTGTAATAAAT	3318
Db	1054	-----CysAsnAsnAspLys-----CysAlaLysCysAspLysGlnLysGlnLysLys	1069
QY	3319	AATGATATTAATAGGATTTATGAGAAACAAATGGAATATATATCAGATTAATACAAAGAA	3378
Db	1070	LysThrLysIleGlnProTrrPrrAlaAspGlnTrrPrrAsnGlnLysAsnLysTrrGlnLle	1089
QY	3379	TTACATGAACAAGCAAAAGTCTGTTAGTAATAGTGGATTTGAAGCTTCGCACTCTGCC	3438
Db	1090	LeuTyrTrrGlnAlaLysIleAlaAlaIleAsnGlyLysThrGlnLysSerThrThrThr	1109
QY	3439	AAAAATCATATAGACAGAGATGTTATGAAATTTTTCGGAATTTATACCAAAATAGT	3498
Db	1110	LysAspAspLysAspLysAsnValIleAspPheLeuGlnLysLeuHisGlnLysAlaSerLyr	1129
QY	3499	GGCAAAAGTATTAAGTGGTACTAGTGAAGAAGT-----	3534
Db	1130	Gly-----ThrArgGlyProProProGlnAlaHisProAspArgArgProArgArg	1146
QY	3535	GCTGCATTTGGTACTACACACAGTATGAAGAAATGTTGGAGCAATCTCCATCATATACAGA	3594
Db	1147	AlaAlaThrSerLysSerAspValTyrGlnThrThrAlaGlyTyrIleHisGlnLysAla	1166
QY	3595	AATTTGATGTTGTCAAGTACCAAAATAGCAATTTTGTGATGAAGAAAAAGATGATGAAGAT	3654

Db 1167 ArgThrArgGluCysLeuGlyGlnAsnValPheCysAsnAsnGly----- 1182
QY 3655 AACGAAATATACCTTTAGAGATTAACACAGACATGATGCTGCTGTGATAA 3714
Db 1183 AsnAsnGluTyrAlaPheSerLeuThrProHisGluTyrLysHisAlaCysLysGly--- 1201
QY 3715 AGTGATGCAAAACCAACAGGCTACAGATTAATAAAGGAGGAGGAGGAGGAGGAT 3774
Db 1202 -----AsnGluAsnLysAlaSerSer----- 1208
QY 3775 ACGAATGTAAACAGATGATATACTTAAGAAACAGATGAGAAACAAAGTACAA 3834
Db 1209 -----ProGluGluLeuGlyArgSerSerPhe 1218
QY 3835 GATTGTCTCCAAAAGAAATAGTAATGATATCCGATTCGCAATGCGAAATATAAAT 3894
Db 1219 AspAspHis----- 1221
QY 3895 TTAGTGAAGACCCGCTGTGTATATGCCCTAGAAACAAAGTATATGCTATTC 3954
Db 1222 -----GlnThrProArgProGluLysAspGluValHis--- 1232
QY 3955 TTGCAAAATGATTAATAAATAAATAATCAATCACAAGTTAATTAAGAAGCTTTC 4014
Db 1232 ----- 1232
QY 4015 ATCAATCTGACAGACAGAAACATCTCTCATGATATTATTAAGTAAAGATGT 4074
Db 1233 -----SerSerGluGluGly 1237
QY 4075 GAAGAAATGAACTCGATTAAGAATTAAGAAGGCAAAATTCCTCCGATTTTGAGA 4134
Db 1238 GlnGluLysAspGluSerGluAspGluGluLysGlu----- 1248
QY 4135 TCATGTTTCACATTTTGAGATTTATAGATTTTATTGGAACAGATATATCAAA 4194
Db 1248 ----- 1248
QY 4195 GGTATGTGAGGAGTAATACTAAAGACAAATAGATCTCTTTTCAAAATGGTGAC 4254
Db 1249 -----GluGluValGluGluValHisAspGlyAlaAsp 1259
QY 4255 CAAAAATCTCTTAATGAGAAACACGCCAAGATGGTGACAGACATAGTCATGATA 4314
Db 1260 GlnLysAla----- 1262
QY 4315 TGGGAAGCTATGCTATGTGACTAGTAATAATGGGGCAAAAGATATTTACCGAA 4374
Db 1262 ----- 1262
QY 4375 AACTACGTTACAAACAGCTCAATTTAGTACAAAGACACACTTTGAGAGATTTGCC 4434
Db 1262 ----- 1262
QY 4435 AAACGACCCAGTTTTCAGATGGCTAACCGAATGGTACGACACTATGTAACAGA 4494
Db 1262 ----- 1262
QY 4495 CAAAAATATTTAGAGATGTGAGAAATTAAGTCAATGACCAATTAAGTGTAT 4554
Db 1262 ----- 1262
QY 4555 ACAGATGTAAATAGAAATGCGAGACTACGTTAAATATATGAAAAAAGAGTGG 4614
Db 1262 ----- 1262
QY 4615 ATTCCACAAGATTAATATTAACAAGATGAGCGCAAAAAAGATTCGATAGACACAC 4674
Db 1262 ----- 1262
QY 4675 ATTGTTAATGTTACAGACTATACGAAAGATGACAGATTAATCTTGAACAGAAA 4734
Db 1262 ----- 1262
QY 4735 TTTACTCTAGTGTGTGTATTAAGCTGAAGAGTCCCTGTGTGTAACAAGAAATATACA 4794
Db 1262 ----- 1262
QY 4795 TTGTAGAAAAACAGGCTTACTATGATGCCAGACAAATTTGTGGTGACAAATTTAAT 4854
Db 1262 ----- 1262
QY 4855 GAAATAGCAGCAAAATATACATTTTGAGTAAGATAGTCAAAAGATTAAGTAAG 4914
Db 1262 ----- 1262
QY 4915 GAGGCAAAACACAGTCTATTAAGTGGCAAAACAAAGGCTTAATATACATTAATTG 4974
Db 1263 -----GlnValAlaSerGlnProGluAlaSerProThr----- 1273
QY 4975 AAAGAAATGACATGATGATGCTTTTCTCTCTGCGACTACGATATGTTTCATGCA 5034
Db 1274 -----ThrLysAspValAla----- 1278
QY 5035 TTGATGCAATTAATACAGATCCAGAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 5094
Db 1278 ----- 1278
QY 5095 ATGGAAGTGGCGCAACGGAAGGTACATTTGGGCAATACTACAAAGAAAAAGAA 5154
Db 1278 ----- 1278
QY 5155 AAAGCAAAATTAACAAAGCTCGATGCGCAAAATATCTTATGAGAGTCCCGCTGTAGT 5214
Db 1279 -----LysProProCysAsp 1283
QY 5215 GCTAAGAAATATAGTTTATATTAAGATATTAATTAAGTATTAATTAATTAAGT 5274
Db 1284 IleValLys-----GluLeuPheSerAsnValAspThrLeuGln 1296
QY 5275 GATGAAAAACAAAGACGAGGAAATTTGAAGAAATATTAACAAAAATGGAACATCA 5334
Db 1296 ----- 1296
QY 5335 GTTGGCAAGAGATGATAGTACTACAGAAATCCCGTACTGCCGAAATTTTTC 5394
Db 1296 ----- 1296
QY 5395 TGAACGAAATTAAGGATGTGTGGAACGATATGATTCGGGTACAAACGTGTAGG 5454
Db 1296 ----- 1296
QY 5455 GATGATGAAATAGTGAATAGTGAAGAAGTGAAGATCTAATAAATGTTGTTCT 5514
Db 1297 -----LysAlaCysSerThr 1301
QY 5515 GTACCTCAGATGATTAATTCCTATGGGAAATATGCCATGAGGTACTGCGTATCAG 5574
Db 1302 -----LysTyrGlyProGlyLysAsn----- 1308
QY 5575 TTTCTTGATGTTGGCGAATGGGTGAAGATTTTTCAAACATAAAGAAAGATTG 5634
Db 1308 ----- 1308
QY 5635 GAGAAATTTGTAAGGGCGGTGAATGATTATATCTTGTGATATAGTAAGTAAGAAAG 5694
Db 1308 ----- 1308
QY 5695 AAATGTACAGATGCTGTACACATATATAAATTTATATAGTGAAGTGAAGACACAGTAT 5754
Db 1308 ----- 1308
QY 5755 GAAAAACAAATCAAAATATGTTGTAAGTAAAGCAAAATATATTCGAGATCTGTG 5814
Db 1309 -----TyrGly----- 1310

QY 5815 GCAAAAGATGCAGACGACGCTCCGGAATATTAGACAAACATTTAAATTTTGCA 5874
 Db 1310 ----- 1310
 QY 5875 AATAAAGTGAGATGTGAATATTAAGTATGAAAGATGTGCACACAGCATTAAT 5934
 Db 1311 ----- 1311
 QY 5935 GATGTAAATAGTCAAAATATGCGGATCATATGACGATGACCAAAAGATTGAAGA 5994
 Db 1324 G1YGLuSpGLyGLn ----- 1328
 QY 5995 AAGTGAATTTGTCAGTCCACAGAGTCCACACGTGTACGAGGCAACACCGTCACCA 6054
 Db 1329 ----- 1339
 QY 6055 CGGGTATCACTGATATCAAAAGGAGCATCGAAAAAAGAGGAAACAGCGCGCT 6114
 Db 1340 ----- 1354
 QY 6115 ACAAAACAGCCGAAAGAAAGTGAATCTTAACAAGAAATGCGAGCACAACAGAAC 6174
 Db 1355 ArgArgArgArgLeuTyrValG1YGLyLeu ----- 1364
 QY 6175 CGACGACGACACACACACACACACACATCAACACACACACACACACACATTCAC 6234
 Db 1365 ----- 1380
 QY 6235 GTGGGCACATGTGTAAAGGCATCTTTCGAATTAACCATAGACATGAGGATGAG 6294
 Db 1381 G1uAlaThr ----- 1387
 QY 6295 GGTGTAAATCCAAAGCATATGACATATCCATAATGGGTTGTATGTAGTAAGTCT 6354
 Db 1388 ----- 1395
 QY 6355 AAGAAATATGAAATGCGATATGATCCCTCCAGAGAAATATATGATATAATAT 6414
 Db 1396 LysGluSerSerGly ----- 1401
 QY 6415 ATCAATATTTAAATTTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 6474
 Db 1402 ----- 1414
 QY 6475 ATTAATATGTCAGCAATAGAACCTAATTTTGTGTTAAATATATATATATAT 6528
 Db 1415 I1eGluSerAlaIaValG1uThrPheLeuTyrPheLeuTyrPheLeuTyrPheLeu 1434
 Db 6528 ----- 6528
 QY 1435 Alag1uLysLysAlaGluLeuGlnAsnG1YLeuAspLeuTyrSerSerG1YAspG1Y 1454
 QY 6529 ----- 6579
 Db 1455 AspProAspAsnPro ----- 1472
 QY 6580 AAAAGAAATATGAT 6639
 Db 1473 LeuArgLeuMetPheTyrThrLeuG1YAspTyrArgPheLeuValHisG1YLys 1492
 QY 6640 TCTAATCATATAAAATATATATATATATATATATATATATATATATATATAT 6699
 Db 1493 ThrSerAspSerG1YAsnThrAsnG1YSerAsnAsnAsnAsnI1eValI1eGluAlaSer 1512
 QY 6700 AATAAGAAAGAAAGATATAAAAGATAGAAATATATATATATATATATATATAT 6747
 Db 1513 G1YAsnLysG1uAspMetG1uLysI1eGlnG1uLysI1eGluG1uLysLeuProLysAsn 1532
 QY 6748 ----- 6759
 Db 1533 G1YG1YThrProLeuValProLysSerSerAlaGlnThrProAspLysTyrTrpAsnGlu 1552
 QY 6760 AATAAATAATTTATTTGGAGAGATCATATATATATATATATATATATATATAT 6804

Db 1553 HisAlaGluSerI1eTyrPheLysG1YMetI1eCysAlaLeuThrTyrThrG1uLysAsnPro 1572
 QY 6805 -----CATCTCACAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 6849
 Db 1573 AspThrSerAlaIaValAspG1uAsnLysI1eGluLysAspAspG1uValTyrGluLys 1592
 QY 6849 ----- 6849
 Db 1593 PhePheG1YSerThrAlaAspLysHisG1YThrAlaSerThrProThrG1YThrTyrLys 1612
 QY 6850 -----CAGTACAT---GACATGACCAACTG----- 6873
 Db 1613 ThrGlnTyrAspTyrG1uLysValLysLeuGlnAspThrSerG1YAlaLysThrProSer 1632
 QY 6874 -----ACGCTCCCTGAGAGAGTTGTAAAGGCGCCCAATTTTGAGATG 6921
 Db 1633 AlaSerSerAspThrProLeuLeuSerAspPheValLeuArgProTyrPheArgTyr 1652
 QY 6922 TTCACAGAAATGGCGAGAAAGATTTGTAAATAGAGAAAGAAAGAAAGAAAGAAAG 6981
 Db 1653 LeuGluGluTyrPheG1YGlnAsnPheCysLysGluArgLysLysArgLeuLysGluI1eLys 1672
 QY 6982 GCGGCTGT-----AACGATATGAGTGAATGCT----- 7011
 Db 1673 GluGluCysMetAspG1YSerAspLysLysTyrSerG1YAspG1YGluGlnCysAspArg 1692
 QY 7012 -----AGTAATGACGGTAAAGACAAAGAAAGAAAGAAAGAAAGAAAGAAAG 7050
 Db 1693 ArgAspThrSerAsnGluValSerAlaAspLeuGluArgSerCysG1YAsnSerCys 1712
 QY 7051 GTAACATATCAAAATTTTATTAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 7110
 Db 1713 ArgPheTyrLysLysTyrPheLysArgLysArgLysGluTyrAspLysGlnAlaAsnAla 1732
 QY 7111 TTCAAAGAAAGATTAAGATGCGAAAGATATATATATATATATATATATATATAT 7170
 Db 1733 TyrSerLysGlnLys-----ThrLysTyrGlu----- 1741
 QY 7171 GAGAAAGCAACATGTGCTCATGATATTTAAACATGAATTAAGAAATATGTGCAAT 7230
 Db 1741 ----- 1741
 QY 7231 AAGGATTTCTTGTATGACAAAGAAAGCTTCTTACACACTACCAAAAGACAAAGATCA 7290
 Db 1742 ----- 1744
 QY 7291 CAATCATCCGATGCTATATGATATGCGAAGATCGCTGATATATGCTCGAAGATTTAAC 7350
 Db 1745 LysG1YAlaG1YLeuAsnAsp-----HisAsn 1753
 QY 7351 AAGTGTAGTGTCTGTAAGCTTCAAAAGAGATCTATGATTCATACAAAAAATTAAT 7410
 Db 1754 LysGluPheCysValLysLeuG1Y----- 1761
 QY 7411 GAACCTAAATATCTATGATATGTTAGAGAAAGCAACCATATATAT----- 7455
 Db 1762 -----ThrCysThrAspAlaIaIaPheLeuAsnArgLeuLysAsn 1775
 QY 7456 -----TTACTTAAGAAGCAAGAAAT---AATATGATTTTACCTTG---AAG 7497
 Db 1776 G1YProCysLysLysAspAsnGlnAsnG1YLysAsnAspI1eAsnPheG1YAsnThrGlu 1795
 QY 7498 GAAATATTTATCTATGATGATCTACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 7557
 Db 1796 GluThrPheArgProAlaGluAsnCysLys----- 1805
 QY 7558 AATATATCTTGGATCTTAAGAAAGCTTATGACCTGATATAATATATATATATATAT 7617
 Db 1806 -----ProCysSerSerPheLys-----I1eAsnCysArgAsn 1816
 QY 7618 ---CCTTGTGAAATATAGAAAGAAATCGTTTAAAGTATGATTTAGATGAATGAATGAT 7674


```

QY 9754 CAACAGCTGTCCTAAGACACTCTATTCTTACACCCATGATGATCTTTCTATCA 9813
    |||||
Db 2158 -----ProlysPro-----Gln 2161
QY 9814 GCACCTTATTCACACATCGAGTAGCACAAATGATGCTTAAATGATATATGAA 9873
    |||
Db 2162 GlnProProAspLeuSerHisProAla-----VallePro 2173
QY 9874 AGTATCTCTCTGTTGTTATGATCGCGCTAGCTTTG-----ATAGCGCTT 9921
    |||
Db 2174 SerLeuValHisSerThrLeuAlaIlePbSerValGlyIleGlyPheAlaIlePheThr 2193
QY 9922 CATTTATGAGAAAAAATTCAAATGCTGTG---GACTTGTTCGTATATGATATC 9978
    |||||
Db 2194 PheThrLeuLysLysLysThrLysSerSerValGlyAsnLeuPheGlnIleLeuGlnIle 2213
QY 9979 CCGCAAGAGAGATGAGAACCTAGCTTGGAATCCAAATAGTTCATACCATATAGA 10038
    |||||
Db 2214 ProLysSerAspLysPheProThrLysLeuSerProAsnArgTyrIleProTyrThr 2233
QY 10039 AGTGTTCATATTAAGGCAAAACATATATATATATGAGAGAGATACT-----AGT 10089
    |||||
Db 2234 SerGlyLysTyrArgGlyLysArgTyrIleTyrLeuGlnLysPheSerGlyThrAspSer 2253
QY 10090 GGAGATGACATATATATATGCTGACTTATCTCTGATATTTACTTATCCGAAAGT 10149
    |||
Db 2254 GlyTyrThrAspHisTyr-----SerAspIleThrSerSerGlnSer 2267
QY 10150 GAGTATGAGAAATGATATTAATGATATATATATGATGAGAGAGATACT-----AGT 10209
    |||||
Db 2268 GlnTyrGlnLysMetAspIleAsnAspIleTyrValProLysSerProLysTyrLysThr 2287
QY 10210 TTGATAGAACTAGTACTAGAACCAAAAAGGATATACCAAGTGTAT----- 10257
    |||||
Db 2288 LeuIleGlnValIleLeuGlnProSerGlyAsnAsnThrThrAlaSerGlyAsnAsnThr 2307
QY 10258 -----GATACACCAACT----- 10269
Db 2308 ThrAlaSerGlyAsnAsnThrThrAlaSerGlyLysAsnThrProSerAspThrGlnAsn 2327
QY 10270 -----AATGATACACACAGCAGCAATGATATTTATGATGATGATGATGATG 10320
    |||||
Db 2328 AspIleGlnAsnAspGlyIleProSerSerLysIleThrAspAsnGlnIlePheGlnIle 2347
QY 10321 AACATGATTTTGTATCTCAATATTTA-----CCAAATACAGCAAAATATATAT 10371
    |||
Db 2348 LysAspGlnPheIleSerGlnTyrLeuGlnSerGlnProAsnThrGlnPro-----Asn 2365
QY 10372 TACAAAAGTCAGATATCCATGATGATACAGAACCTAATCTTATATCTGATATCCT 10431
    |||||
Db 2366 MetLeuGlyTyrAsnValAspAsnAsnThrHisProThrThrSerHis-----HisAsnVal 2384
QY 10432 GAAGAAAAACCTTTATATATATATATATATGATGATGATGATGATGATGATGAT 10491
    |||||
Db 2385 GlnLysPheProPheIleMetSerIleHisAspArgAsnLeuPheSerGlyGlnIleTyr 2404
QY 10492 AGTTAT-----AATATTAATATAGTACTAAT-- 10518
    |||||
Db 2405 AsnTyrAspMetPheAsnSerGlyAsnAsnProIleAsnIleSerAspSerThrAsnSer 2424
QY 10519 -----ACTAATATATGATATCCATGATGATGATGATGATGATGATGATGAT 10563
    |||||
Db 2425 MetAspSerLeuThrSerAsnAsnHisSerProTyrAsnAspLysAsnAspLeuTyrSer 2444
QY 10564 GGTATATATTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10623
    |||||
Db 2445 GlyIleAspLeuIleAsnAspAlaLeu--SerGlyAsnHisIleAspIleTyrAspLeu 2464
QY 10624 TATTTG 10628
    |||||
Db 2464 etLeu 2465
RESULT 8

```

```

T28157
erythrocyte membrane protein 1 - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T28157
R:Yang, S.Q.; Wollish, W.S.; Gut, J.; Wu, J.; Ahn, J.; Petersen, C.; Fujioaka, H.; Aik
submitted to the EMBL Data Library, July 1995
A:Description: The molecular cloning and DNA sequence analysis of Plasmodium falcipar
A:Reference number: 220479
A:Accession: T28157
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2212 <YAN>
A:Cross-references: EMBL:U01083; NID:g1517813; PID:g1517814; PIDN:AAB06961.1
C:Genetics:
A:Gene: EMP1

Alignment Scores:
Pred. No.: 2,22e-124 Length: 2212
Score: 2678.50 Matches: 719
Percent Similarity: 44.45% Conservative: 311
Best Local Similarity: 31.03% Mismatches: 809
Query Match: 13.77% Indels: 478
Gaps: 84

US-10-087-013-1 (1-10628) x T28157 (1-2212)
QY 91 AGTCACAAAGTCGCAAGAAATGTTTGAACGTTATGCAAAATATATAGACATCATCA 150
    |||||
Db 9 ThrAsnLysSerAlaLysGlnValLeuAspGlnIleGlyGlnThrIleGln----- 25
QY 151 AATATATGCAAAAGACATGCGATTCG-----TTGAAAGGGGATTTG 192
    |||
Db 26 -----LysLysAlaHisSerAspAlaAspThrPheArgSerGlnLeuLysGlyAsnPhe 43
QY 193 ACGAAACGACATTTCTGCGTGT-----CCTTACGCGCGATTAATAG 237
    |||||
Db 44 GlyIleAlaLysPheTyrAsnGlyGlyIleMetGlnProAsnSerLysLeu----- 61
QY 238 CATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 297
    |||||
Db 62 -----CysGlnLeuAspHisThrIleAspThrAsnValThr 73
QY 298 TATGATATGATGATTTGACATCTCTGCGATGAGACAAACCGATTTGATGAA 357
    |||
Db 74 AspGlyHisSerAsn-----ProCysGlnIleTyrGlnThrValArgPheProAsp 90
QY 358 GATGACAAATCTGAAATG-----GGAATTAATATAGTATTTAAAGAAAAATGATGCT 414
    |||
Db 91 AspAsnArgSerGlnCysThrLysAsnArgIleLysAsp---SerValAspAsnSerVal 109
QY 415 ATAGCCTGTGCGCCACCTAGAACGACATATGCTGATATAAAACCTTGGAAGCTTAAT 474
    |||||
Db 110 GlyAlaCysAlaProTyrArgArgLeuHisLeuGlySerHisAsnLeuGlnSerIleGln 129
QY 475 GATATTAAT-----ACCAAAATATTCATGATTTATGGAATATGATCTAGTACA 525
    |||
Db 130 ThrAsnAsnTyrAspSerSerLysAlaLysHisAsnLeuAlaGlnValCysTyrAla 149
QY 526 GCAAAATACGAAAGGTGATCAATGTTAATATCATCA-----CAATAAGAACT 576
    |||||
Db 150 AlaLysPheGlnGlyGlnSerIleValLysAsnTyrGlnIleGlnLysHisThrThr 169
QY 577 TCAGACGCTGTACTGCTCTTGCAAGAGTTTGCATATATAGCTGATATGATGATGATGAT 636
    |||||
Db 170 GlnClyIleCysThrAlaLeuAlaIleArgSerPheAlaAspIleGlyAspIleIleArgGly 189
QY 637 ATAGATATGTTTAAACCAATGTCCATGAC-----AAAGTAGAAACGGGT 681
    |||||
Db 190 LysAspLeuTyrIleGlnGlyAsnProGlnIleSerAlaArgArgLysGlnLeuGlnAspAsn 209
QY 682 CTCGAGAGGTTTTCAGAAATATACAT----- 708

```

```
Db 210 LeuAdgLySIIlePhaSpHeIIeTyLySgIuLeuThrSerSerArgSngLySthr 229
OY 709 GATGGAATGGAAGATGAATGGAATGATTAACATCCGTGATGATGGAATATATAT 768
Db 220 AsnGlyAlaIngIuArgTyLyS-----AspGlySerGlyAsnTyLyS 244
OY 769 AAATTAAGAGAAGCGTGTGATGATGATGGAATGGAATGGAATGGAATGGAATG 828
Db 245 LysLeuAdgGluAspTrpTrpAsnAlaAsnArgLeuAspTrpTrpAlaMetIleCys 264
OY 829 GATGATCATATTAATCATGATATTTATGATGATGATGATGATGATGATGATGAT 888
Db 265 LysAlaProGlyAsnAlaIngIuTyPhaArgSngThrCysSerAsnGlyGluTySPro 284
OY 889 AATGCTTAATGCGGCGCATTAAGAGAAGGCTTCTCAATTTGATTTGTCCTCA 948
Db 285 GlyGluTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIu 304
OY 949 TATTAGCTGTTGTCGAGATGGAAGAGTGTGCGGAGAGAGAGAGAGAGAGAGAG 1008
Db 305 TyLeuAdgTrpTrpHeGlyIngIuTrpAlaIngIuPhaCysArgTySAsnLeuTyS 324
OY 1009 AAAAAAGTCAAGAGACTGCTGTCTG-----AATGACAAGAGAGCTTATATTT 1059
Db 325 GluAsnAlaIleTySAsnCysHISgIuMetAspAspArgTySgIuTySgIuTyS 344
OY 1060 CATTAATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1119
Db 345 ArgAsnGlyTyTrpAspCysThrTySgIuTySgIuTySgIuTySgIuTySgIuTy 364
OY 1120 AAGTGTACTGCTGTTGTCGATGATGATGATGATGATGATGATGATGATGATGAT 1179
Db 365 GlyCysThrTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 384
OY 1180 GAAGCATTTTAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1239
Db 385 LysGluPhaHeGlyIngIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIu 404
OY 1240 GATTAACAATTTTGTCAATATATATATGATGATGATGATGATGATGATGATGAT 1299
Db 405 SerGluAsnSerProTySAsnTyTrpAsnAsnMetTyGluThrAspHeTyGlyAsn 424
OY 1300 AAGGAAGCGCATATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1359
Db 425 LysTySAspTyGluInsPheMetAsnAsp---PheLeuTySgIuTySgIuTySgIu 443
OY 1360 TGTAAAGAGAGATTAACAGAGAGAGAGATTAATTAATTAATTAATTAATTAAT 1419
Db 444 CysThrAsnIleLeuAspGlyTySgIuTySgIuTySgIuTySgIuTySgIuTyS 461
OY 1420 GGGATATTTTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1473
Db 462 GluThrPheSerHisThrGlyTySgIuTySgIuTySgIuTySgIuTySgIuTyS 481
OY 1474 TGTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1533
Db 482 AlaAspGly---ThrTySAspTyGlyTySgIuTySgIuTySgIuTySgIuTySgIu 500
OY 1534 GACTATTAACCTCATGGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1593
Db 501 LysTyGluProTyGluGlyTySgIuTySgIuTySgIuTySgIuTySgIuTySgIu 520
OY 1594 GAACAAGGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1653
Db 521 GluAsnTySAspIleIleValTySgIuTySgIuTySgIuTySgIuTySgIuTyS 540
OY 1654 GATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1713
Db 541 PheTyS---AsnGlyGlyTySgIuTySgIuTySgIuTySgIuTySgIuTySgIu 558
OY 1714 CTGAACAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1773
Db 559 LeuGluAsnGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 577

OY 1774 TTTGATTAATGCTTACATTTTATTAAGGATGATTAATGATGATGATGATGAT 1833
Db 578 LeuMetPheTrpAlaIleAsnMetLeuTySAspSerIleGlyTrpArgSerTyS 597
OY 1834 ACTGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1893
Db 598 AsnCysLeuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 617
OY 1894 TTTGACAGATGAGTTAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1953
Db 618 TyGluTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIu 637
OY 1954 AAAAAAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2010
Db 638 LysGluThrAspHeGlyIngIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 657
OY 2011 ---TATTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2067
Db 658 AspGluPhaHeTrpTrpAspTrpTrpAspTrpTrpAspTrpTrpAspTrpTrp 677
OY 2068 ATGGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2127
Db 678 GlyGluMetLeuGlyTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 695
OY 2128 TTAGAGATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2184
Db 696 MetTySgIuThrIleLeuAspGlyTySgIuTySgIuTySgIuTySgIuTySgIuTyS 715
OY 2185 -----AAGACAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2226
Db 716 ValAsnHisTySAspAsnAsnTySgIuTySgIuTySgIuTySgIuTySgIuTyS 735
OY 2227 -----GCACAACAACCCGCTGTGATTAATTAATTAATTAATTAATTAATTAAT 2277
Db 736 AspIleProGlyAlaGlyAsnProCysAlaTySgIuTySgIuTySgIuTySgIuTyS 755
OY 2278 AATTAATAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2337
Db 756 ---ValAsnTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 774
OY 2338 GGTCTT---CATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2394
Db 775 GlyValSerTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 794
OY 2395 AGAAGAGCTTCAAGAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2454
Db 795 ThrIleTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIu 813
OY 2455 GGTTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2514
Db 814 GlyHisThrAspGlyProCysTySgIuTySgIuTySgIuTySgIuTySgIuTySgIu 830
OY 2515 GTCTAGGAATGAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2574
Db 831 ArgThrGluAspGlyTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 849
OY 2575 ATTATGCTCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2634
Db 850 TyMetProProArgGlyGlyHisPheCysThrSerAsnLeuGlyTySgIuTySgIuTyS 869
OY 2635 GATCACCCTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2685
Db 870 AsnTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 889
OY 2686 TTGGGAGATGTTTCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2745
Db 890 LeuGlyAspValLeuPheAlaIleAsnTyGlyAlaAspHeIleTySgIuTySgIuTySgIuTyS 909
OY 2746 GAAAGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2805
Db 910 LysGluAsn-----AspTyTySgIuTySgIuTySgIuTySgIuTySgIuTySgIuTyS 920
```


QY 2806 TGTCCGCTATACGTTACAGTTTTCAGATATAGTGATATATATTCAGGAGAGATCTC 2865
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 921 CysAlaIleMetLysTyrSerPheAlaAspLeuGlyAspIleLeuGlyThrAspMet 940
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 2866 TGGGAGAGAAACGGTGACATGTGTAAGCTGCAGAGACATTTGGAACTGGTTTGGTAAT 2925
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 941 TTPAsAsnAspSerLysGluSerLysThrArgAspLysLeuArgGluIlePheAspThr 960
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 2926 ATACATTAAGTCACTCAAGCAAGCAAAAGAAATGATTAATTAATGATGATGCCCAATAT 2985
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 961 IleLysLysHisProGlyLeu---LysGluIleTyrLysGluAspThrPro---Tyr 978
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 2986 TTTAAATTAAGGAAATATGGTGCGAGAGCTAATAGCCAAAGATAGGAGCAATGAAA 3045
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 979 ThrLysLeuArgGluAspThrProGluAlaAsnArgLysLysIleThrGluAlaMetGln 998
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3046 TGTGATATATAATATTGAAGATAATCGGACCAATCAACCAAGTATATTCG 3105
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 999 Cys-----ProThrProAsnGlySerPhePro 1007
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3106 GGATATATGATCATACACCATTTGATGATATATATCCCAAAATTAAGATGATGACC 3165
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1008 CysLysSerTyrHisThrProLeuAspAspTyrIleProGluArgLeuArgTrpMetThr 1027
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3166 GAATGGCGAGATGTACTTCGACAGGTGCAGAAAAGAGATATGATTAAGTGAAGGAGAG 3225
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1028 GlutProAlaGluThrPheCysLysGluGlnLysGlnTyrGluIleuValSerAla 1047
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3226 TGTAAAGAGTGTAAAGATTAAGTGAAGCTGTACGAAAGAGTGTGTACAGGT 3285
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1048 CysAsnGlyCysLysAspGlu-----GlyLysValCysThrAsnLysSer---SerGln 1064
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3286 TGTACAGAGTGCACAGACCTTGTATGAATATATGAATATATGAATATATGAATATGA 3345
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1065 CysThrSerCysMetGlnAlaCysGluAsnTyrLysAsnPheIleAsnThrTrpLysGlu 1084
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3346 CAATGGAATATATATACAGATTAATACAAAGATTTACATGACACACCAAAATGCTGT 3405
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1085 GlnThrAspLysMetGluIleLysTyrLysLeuTyrLeuGlnIleGlnThrThrAla 1104
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3406 AGTAAATAGTGTAAAGCTTCCAGCTACTGACCCAAATATATATACAGAGATGTAAAT 3465
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1105 AlaAsnGlyGlyProAspThrTyrSerGlyLeuValAspGluAsnGluLysProValAla 1124
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3466 GAATTTTGTCCGATATATACCAAAATGTCGCAA----- 3504
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1125 AsnPheLeuPheGluLeuTyrLysGluAsnGlyGlySerIleGlyAsnProArgAspThr 1144
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3505 -----AGTAATTAAGTGTACTACTGATGAAAGCTGCTACTTGTGTACTACACACAG 3558
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1145 ProArgAlaLysArgSerLysArgGluThrAlaProAlaSerValAlaLysAsnAspVal 1164
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3559 TATGAAAAATGTTGGAGCATATCTCAT-----GATACAGAAATTTTATGATGTCTCATGCA 3615
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1165 TyrSerThrAlaAlaGlyLysValHisGlnGluMetGlyProHisMetGluCysLysThr 1184
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3616 CAAATAGCTTTTGTGATGAGAAATGATGTAAGATTAAGCAAAATATGCTTTTGA 3675
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1185 GlnThrGluPheCys---GluLysThrAspGluGlnTyrAsnGluAsnTyrThrPheLys 1203
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3676 GATTAACACAGAGCATGATGT 3735
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1204 AsnProProGluGlnTyrLysAspAlaCysIleCysAsnThrArgProProLysGlu 1223
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3736 GTACAGATTAATAAGAAAAA-----AAAGCG 3762
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1224 AspSerArgLysArgSerGluAspSerAspGluGluLysValLysGluThrLysVal 1243
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3763 GAAGAAAAAGATACGAA----- 3780
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1244 GluGluLysAlaThrGluAspAlaValAspThrGlyProProProAlaProLysGluAla 1263
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3781 -----GTAAACAGTGAATGATATATCTTAAAGAAAAAGATGGAAG 3822

DB 1264 ThrThrThrLeuAspValCysProIleValAlaGlyValLeuThr-----Lys 1279
QY 3823 AAACAGTAAAGATTTGTCAATCCAAAAAGAAATGAAATGATATCCC---GATTGGCA 3879
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1280 GluAsnLeuGluAsnAlaCysProThrLysTyrGlyProLysAlaProThrSerTrpLys 1299
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3880 TGC-----GGAAATATA 3891
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1300 CysIleProThrGluLysThrAsnAlaAlaThrGlySerGluLysSerSerGlyAsnGly 1319
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3892 AATTAGTGAAGACCTGCT----- 3912
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1320 AlaLeuGlnArgAlaLysArgAlaThrValGluSerGlySerProValThrSerAsnSer 1339
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3913 -----GTGTGTATGCCCTTACAGACAAAGATTAATGCTA-----CATTTCTTG 3957
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1340 GlySerIleCysIleProProArgArgArgArgLeuTyrIleGlnLysLeuHisAspTrp 1359
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3958 GCAATGATATGAAATTAATAAATAATTAACAATCAAGTT----- 3996
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1360 AlaSerGlyAsnThrValValSerGlyGlnAlaGlnThrProGlnGlyThrSer 1379
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 3997 -----AATTAAAGAGCTTTCAATCAATCTGCAGCA 4029
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1380 ProSerGlyLysGluThrProSerAspLysLeuArgThrAlaPheIleGlnSerAlaAla 1399
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4030 GCAGAAACATTTCTTCATCGTATTAATTAATAAGTAAGTGTGAAGAAATGAACTG 4089
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1400 IleGlnThrPhePheLeuThrPheAspArgTyrLysLysGlu-----LysGluIle 1415
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4090 GATTAAGATTAATAAGAA----- 4107
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1416 GluLysLysGluLysLysValAlaAsnGlyGlyLeuValProSerLeuAsnGlyLysPro 1435
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4108 -----GGCAAA 4113
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1436 ProGlnGlnProGlyValThrGlyAspSerProGlnSerLysLeuGlnGlnThrGlyVal 1455
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4114 ATTCTCCCGCATTTTGGAGATCCATGTTCTTACACATTTGAGATTAATAGATTTTWTAA 4173
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1456 IleProProPheLeuArgGlnMetPheTyrThrLeuGlyAspTyrAlaAspIlePhe 1475
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4174 TTTGGAACAGAT-----ATATCAAAAAGTCAATGCTGAG----- 4206
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1476 PheGlyLysAsnAspIleValIleAspThrLysAsnGlyAspLysAspIleAlaGluArg 1495
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4207 GGAAGTAATTAATAAGACAAATAGATTCCTTTTCAAAAATGTCGACCAAAATCTCCT 4266
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1496 GluLysLysIleLysAspAlaIleGluArgValLeuLysAsnAlaAspSerGlnProPro 1515
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4267 AATGAAAAACAGCCAGAAATGTGACAGACAGATAGATAGATAGATAGATAGATAGATAG 4326
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1516 SerAspGluLysArgGlnThrTrpTrpGluGlnAsnGlyLysIleThrPheAsnGlyMet 1535
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4327 CTATGTGCTACTAGTA-----AAATTT 4347
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1536 IleCysAlaLeuLeuTyrLysGluLysAspGluLysGlyThrProLeuLysGlnAsnGlu 1555
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4348 GGGGCAAAA-----AAAGATATTTTACCAGAAATCTACGCT 4383
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1556 GlyLeuLysSerAlaLeuThrPheAspGluLysAsnLysSerProLysAspGlnLysTyrGln 1575
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4384 TACACACAGTCAAAATTTAGTCAAAAGC----- 4413
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1576 TyrAspLysValLysLeuAspGluAsnSerGlyThrSerProLysThrAsnAspHisVal 1595
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4414 -----ACACTTGGAGGATTTGCCAAACAGACCCAGTTTACAGATGCTAACCGAA 4467
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 1596 ProProThrProThrAsnPheIleSerArgProProTyrPheArgThrLeuGluGln 1615
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
QY 4468 TGTAGACAGCATATGCTATACAGCAAAATATTGTAAGATGTGCGAAAAATGT 4527
DB ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :

Db 1616 TrpGluGluThrPheCysArgGluArgLysLysArgLeuGluLysLysValGluCys 1635
 QY 4528 AAGTCAAAATGAC-----CAATTGAAGTGTAT----- 4554
 Db 1636 MetAspGluAspGlyLysLysGlnLysCysSerGlyAspGlyLysGluLysLeu 1655
 QY 4555 -----ACAGATGTAATAGAAATGC 4575
 Db 1656 ArgLysGlnAspLysSerThrValArgAspPheTyrCysProGluCysGlyLysTyrCys 1675
 QY 4576 GAGACTACGTTAAATATATGAAAAA-----GAGTGGATTCCACAAGATTAATAT 4632
 Db 1676 ArgPheTyrLysArgTyrPheGlnLysLysAspLysAspLysGlnLysGlnLys 1695
 QY 4633 TACAGAGTACGCGCAACAAAAAAGTTCGATGACACACACATGGTGAATGTTACA 4692
 Db 1696 TyrAsnGlnLysThrAspAlaArgArgAsnAsnAsnAspAlaPheSerThrThr 1715
 QY 4693 GACTATACGAGCAACGATTCATCTGAAACGAAATTTACTGCTAGTGT--- 4749
 Db 1716 LeuAspThrCysThrThrAlaGlyAspPheLeuGlnThrLeuLysAsnGlyProCysLys 1735
 QY 4750 -----GGTAAAGCTGGAAGTCCCTGTGTGTACAAAGAAATATACAATG 4797
 Db 1736 AsnAspAsnValAspAspSerGlyGluAsnLysLysLysPheAspGluAsnGlyAspThr 1755
 QY 4798 TTAGAAAAACAGGCTTCTATGATGCGACAAACATTTGGTGCACAAATTTATTGAA 4857
 Db 1756 PheLysTyrThrGlnTyrCysGlyThr-----CysSerLeuAsnGlyPheLysCys 1772
 QY 4858 AATGACACACAA-----TATACTAATTCGAGTAAAGATTAAGTAAAGGATTA 4908
 Db 1773 AsnGlyAspAspCysArgValArgThrAsnValThr-----CysAsnGlySer 1788
 QY 4909 GTAAGAGAGCAACACAGGTCTATTAGTGCACAAAGAGTCTTAATATCAAT 4968
 Db 1789 AsnArgThrThrThrThrLeuThrAlaAspAspLysLysAsnGlyLys-----AsnSerAla 1806
 QY 4969 AACTTAAGAAATGACGTAGAGATGCTTTTCTCTCTGTCGACATGATATGTTT 5028
 Db 1807 GluIleAsnMetLeuValSerAsp----- 1814
 QY 5029 CATGCATTCGATGCAATATATACAGATCCAGAACTTAAAGATGAAAGTGGTCCGAAA 5088
 Db 1815 -----AspIleAsnSerGlyAsnGlyPheAsn--- 1823
 QY 5089 AGATTGATGAGTGGCGCAGCAGAGGCTACAAATTTGGTGCATACACAAAGAAAA 5148
 Db 1824 -----AspLeuGlnLysLysAsnAlaAsnIlePheLysGlyLysLysGln 1840
 QY 5149 AAGAAAAAGAGAAAAATAAAAAGTCGATGCGCACAAATATCTTATGAGGTCCCGCT 5208
 Db 1841 LysTyrLys-----CysValTyrPheCysLysSerAspAla 1852
 QY 5209 TGTAGTGCATGAAATATAGTTTTATGATTTAAGACATATATCTA----- 5256
 Db 1853 CysGlyLeuLysLysAsnAsnAspLysGlnAsnGlnIleIleIleIleArgAlaLeu 1872
 QY 5257 -----GGTATGATTAATTTGGAAGATGAAAAACAAAGACCGAGAAATTTGAG 5307
 Db 1873 PheLysArgTyrPheGlnLysLysPheLeuAspAspLysLysLysLysLeuAsn 1892
 QY 5308 AAAAATTTTACAAAAATGGAACATCACTGGCAAAAGAGATGATAGTACAGAAAT 5367
 Db 1893 ProCysLysAsn-----GlyGlnLysAlaIleCysThrAsnGlyLys 1907
 QY 5368 CCCGCTAGTACTGCGCAAAATTTTCTGGAACGAAATATGAGATGCTGTGCAACCA 5427
 Db 1908 ValGluGln-----TyrLysGlnLysArgThrGlnTyrThr--- 1920
 QY 5428 ATGATATCGGGTACAAAGTGTAGGATGATGAAATATGAGAAATAGTCAAGAGT 5487
 Db 1921 -----AsnIleLysAsn-----ArgPhe 1926

QY 5488 GATGAGATCTATAAAAAATGCTGTTCTGACCTTCAAGATGATTAATCTATGGGAAA 5547
 Db 1927 AsnGlnLys-----AsnGlyAspAspPheGlnLys 1938
 QY 5548 AATCGGATGAAAGTACTGCTATCATCTTCTGATGTTGGCCGAATGGGTGAAGT 5607
 Db 1939 SerSerPheArgSerPheLeuValAspLeuIleArgGlnIleAla-----AlaThr 1955
 QY 5608 TTTTGAACATTAACAAAAAGCAATTTGGAATTTGGAGGCG-----TGTAAATGAT 5661
 Db 1956 IleAspLysGlyAsnIleAsnGlyLeuValLysLeuValLysSerValLysCysAsn 1974
 QY 5662 TATACTGTGTCATATATGATGATTAAGAAAGAAATGACAGATCGGTACACAAAT 5721
 Db 1975 -----CysGlyAsnAsnSerGlnAsnGlyLysGlu----- 1984
 QY 5722 AAAAAATTTATAGTACGAGGAACCCACATGATGAAAAAATCAAAAAATATGTTAG 5781
 Db 1985 -----GlyGln 1986
 QY 5782 AATTAAGCAAAATATATTCGACATCTGTGCAAAAGATGACAGAGCTGCGAA 5841
 Db 1987 GluAsnAspLeuValLeuLys-----LeuLeuGln 1996
 QY 5842 TATTTAGCAAAATATTAATAAATTTGTGAAAT-----AAAGTGAATTTGGA 5895
 Db 1997 LysLeuGlnLysLysAlaGlnLysCysLysAspAsnProGluThrSerGlyIleProGln 2016
 QY 5896 TATTAAGTATGAAAGATGTGTCCACACAGCATTAATGAT----- 5937
 Db 2017 GlnProCys-----GluValSerProAsnIleGlnAspGlnGlnProLeuGln 2034
 QY 5938 -----GGTAAATAGTCAAAATATATGCCCCCATCATTAAGCAT-----GAACCA 5979
 Db 2035 GluGlnGluAsnThrValGlnHisProLysIleCysAspAspValLeuLysThrGlnPro 2054
 QY 5980 AAGAGATTTGAAGGAAATGTAAATTTGCACTGCCACAGGTCACACAGTGAACAGG 6039
 Db 2055 GlnProGlnGlnProGlnGlnLysGlnLysSerProGlnProLysAspValLysGln 2074
 QY 6040 GAAACCCGTCACACGCGGTATCACTGATATCAAAAGCGACGCGATGAAAAAGAACCG 6099
 Db 2075 GluGlnGlnLysGlnLys-----LysGlnGlnLysAspLysGlnGln 2089
 QY 6100 AAAACGCGCGCTTACAAAAACCGCAAAAGTGAATAATCAACAGAAATGCGA 6159
 Db 2090 ProGlnLeuProProProThrProProAlaIleProSer-----SerThrProIle-Pro 2108
 QY 6160 GCACAAACACGACCCGACGACGACGACACAAACACGAAACGACAT 6208
 Db 2108 oProLysProLysProProProGlnValGlnLysAsnProTyrLys 2124

RESULT 9
 T14602
 Variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragment)
 C:Species: Plasmodium falciparum
 C:Date: 20-Sep-1999 #Sequence: Revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14602
 R:Yoss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
 A:Submitted to the EMBL Data Library, February 1998
 A:Description: Identification of a conserved 5' flanking region of Plasmodium falcipara
 A:Reference number: Z18158
 A:Accession: T14602
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2135 <VO>
 A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AA05220.1
 A:Gene: varpi17

Alignment Scores: 6.29e-106 Length: 2135
 Pred. No.:

Score:	2303.00	Matches:	681
Percent Similarity:	42.62%	Conservative:	306
Best Local Similarity:	29.40%	Mismatches:	745
Query Match:	11.84%	Indels:	584
DB:	2	Gaps:	95

US-10-087-013-1 (1-10628) x T14602 (1-2135)

OY	61	GGAGCTGCTAAGACCCCTATTATTAAGAAGTCAAAAGTCGAAGAAATGTTTGGA	120	
Dd	5	GlyAspProGlnGlyGlyThrLysAsp----	AspIleLysHisMetHeAsp	22
OY	121	CGTTATGCCAAAATA-----AGACAATCATCAAAATATGCAAAAGACATGTGCATTG	177	
Dd	23	ArgIleGlyThrIleGlnLysGlnValHisAsnAlaIleGlnGluPheArgAsnIu	42	
OY	178	TTGAAAGGGATTGACGAAAGCAGAAATTTCTGTGTCTTCAGCCACTATAAG	237	
Dd	43	LengInglyArgLeuSerSeraIa-----ThrsnIleAsnLeu	55	
Dd	238	CATATATATTATTATTCATATCCATGTGAATTTAGATCATAG-----GAACATCATAT	291	
OY	56	LysLeuIleGlyAsnIleGluThrCysSerLeuGluTyLysTyTLysHisProAsn	75	
Dd	292	TTCAGTAGTCATGTATGAAATTTGACACATCTTGCSCATGTAGAGACAAACCGATTG	351	
Dd	76	GlyGlyLysArgValSerAspLysArgTyProCys-----LysGluLeuSerArgLys	93	
OY	352	GATGAACATGMAAGATCGA----TGCGAAATAAAAATKSGTATTTAAAGAAAAAT	408	
Dd	94	TyrValAspArgPheSerAspLysIleGlyGlnCysThrAsnIuLysMetArgArg	113	
OY	409	GATGCTGTA---GCGCTGGCCACCTAGACAGCAGCATGTGTGATTAACCTGGGA	465	
Dd	114	AspGlyIleGlyAlaCysAlaProTyArgArgLeuHisLeuCysHisHisAsnLeuIu	133	
OY	466	GCCTGA-----AATGATATATAATACCACAAATATTT--CATGATTATGGGAATGTA	516	
Dd	134	ThrIleGluThrAsnAsnTyGluSerAsnAsnAlaLysHisAsnLeuValAspAl	153	
OY	517	CTACTTACGMAAATKCGAAGGTGATCAAT-----GTAAATATATATCCAAAT	567	
Dd	154	CysMetAlaAlaLysTyGlnGlyAspSerIleLysThrTyTLyGelLysHisnHis	173	
OY	568	AAA---GGAACCTTAGACGGTGTACTGGCTTCGACGAAAGTTTTGCAGATTAGTGAT	624	
Dd	174	LysTyLysAspSerGlnLeuCysThrValLeuAlaIleSerPheAlaAspIleLysp	193	
OY	635	AATGTAAAGCAATAGATATGTTTAAACCAATGTCCATGACAAA-----	669	
Dd	194	IleValIArgGlyLysAsnLeuPheTyGlyAsnThrGlnIuGlnGluLysArgLinsp	213	
OY	670	GTAAGAACGGGTCTCCAGAGAGGTTTCAGAAATAATCATGATGGAATGGAATGAACTA	729	
Dd	214	LeuGlnAsnLysLeuLysAspIlePheLysIleHisSerGlyLeuThrGlyAlaI	233	
OY	730	AAAAATGATTCATCTCTGATGATGCTGCAAAATTTATTAATTAAGACAGCATGGGG	789	
Dd	234	LysAlaIArgTyLys-----AspThrThrAsnPheTyGlnLeuArgLinspPrtpR	251	
OY	790	AATGTGAATGAAATAAGATAGGGAAGCTATTAACATGTGATGCATCATTAATCTGGA	849	
Dd	252	ThrAlaAsnArgGluThrValTrpLysAlaIleThrCysAspAla-----HisGlyThr	269	
OY	850	TATTTTATG-----CAATCGAAGATGAATACACCATATATTCCAATCTCT	894	
Dd	270	TyrPheArgAlaThrCysGlyLysAspAsnGluSerProSerMetAlaLysAsnAsnCys	289	
OY	895	AAATGCGCCATTAACAAGAGAA-----GTTCCTACCAATTTAATATATGTC	942	
Dd	290	ArgCysGlnLysLysAspGlyArgProAspArgGlnValProThrTyLysPheAspTryAl	309	
OY	943	CCCTCAATATTTACCTGGTTCGACGAATGGGGAAGAGCTTTGGCCGAAAAAGAAATAT	1007	

Dd	310	ProGlnIhrLeuAAtgTtrPheGlnGluTtrPalaGluAspPheCysArgLeuArgGlyHis	329
Qy	1003	AAATGAAAAAGGTCAGAGACCTCTCGT-----AATGACAAAGACGCTTAAT	10533
Dd	330	LysLeuYsaAspAlaIleGlnIlyScysArgGluYsnIlyGluYAlaLeuIlySerLeuYr	349
Qy	1054	TGTAGTCATTAATGGACATGATGTGGACACACATTTTGGAAAAAGATATTTTGATTTG	11133
Dd	350	CysAspLeuAsnArgTyrAspCysGluGlnThrAlaSerGlyLysHisAspPheGlu	369
Qy	1114	GATATTAAGTCTACTGACCTGTGCAGTAATGCAAAAGTTTGGAGTTTGCTAGGACAT	11733
Dd	370	GluAspAspCysIlySgIlyCysGlnTyrSerCysAlaArgPheValAsnTrpIleAspSn	389
Qy	1174	CACAGACGACATTTAAACAAACAAAGAAATATGAAAAAGAAATACATCATATTTA	12333
Dd	390	GlnIlySlySgIupIhrLeuIlySgIlnIlySlyTyGlyThrGluIle---SerLeuYs	408
Qy	1234	TCGACGCTTAACAAATTTGTCATTAATTAATAGCAATATTA-----AAA	1281
Dd	409	SerArgIlySlyAspArgPalaIlyGlyIleSerThrLysValTyrAspGlyTyrGluYs	428
Qy	1282	CAATTTTATGAAAACTTAAGAAAGCAATATGCAATAGACACTTTTATTAATTTA	1341
Dd	429	LysPheYrGluIlySleuYs---SerGluTyrArgThrValGlyGluIupheGluYleu	447
Qy	1342	CTAAATGAGAAAGTATGT-----AAGAGAGATTACAGACAGAAAAAGATATT	1392
Dd	448	LeuSnAsnGluIlySthrCysLysGluValIlyAspAspLysGluGlyIleAspPhe	467
Qy	1393	ACTTTTACTACACTGCGTGAATGACAAAGG-----ATATTTAATCGTCA	1437
Dd	468	LysThrValAsnSerGlySerAlaSerGlyAspAspValAsnIlySthrPheYrArgThr	487
Qy	1438	GAATTTTCCCAAGTGTGTCCGACCTGGCGGCTCAAAATGTGATGAGTAAATACACACAC	1497
Dd	488	LysTyrCysGluIlyAcysProTtrPcysGlyAlaGluGlnIupArgsnIlyValGlyTrp	507
Qy	1498	AAATCAGATTAATGCTGCAACGCTGAATTAATGAAGACATTAACCTCCATGCGGTGTG	1557
Dd	508	LysAlaIlySAspAspArgAspCysSerProGlyAsnAspThrIlyLys-----Tyr	524
Qy	1558	AAGCCTACTAATATCAGTGCCTTTATAGCGTAATGACAAAGCGATATTACACAAAA	1617
Dd	535	LysIlySlySgIupIleThrIleLeuThrGlyAspLysThrLysSerGluIleValGluArg	544
Qy	1618	TTAGAAAATTTTTGT-----	1632
Dd	545	TyrArgIlySphCysIlySAsnAsnGlyLysAsnGlyAlaAsnGlyArgGluGlyIlyAl	564
Qy	1633	-----AACAGCTCAACTTAATTACAAAGAT	1656
Dd	565	GlyIlySerGluAsnGlyAlaAlaSerAsnSerAspAspAlaThrIleCyltyrCysGly	584
Qy	1657	AAAAATATAT-----CAAAATGCGAATGCGATAT-----AAGGAT	1692
Dd	585	GlyGlyAsnSerAspSerSerLeuCysGluIlyTrpThrCysArgTyrTyrIlySlySlyS	604
Qy	1693	GAAAT-----ATTAATAGATGTAACTGAACTGAAACAAATACCTGAAATCAT	1737
Dd	605	GluAsnAsnAspGlyLysAspIleAsnPheCys---ValGlnGlyAlaTrpIlnAsnSer	623
Qy	1738	AATGATATCTAGATTAATATCATCTTCATTAATTTTGGATTAATGCGTTCACATATTTA	1797
Dd	624	LysIlySAspGlnIlySValIlySsetTyrAsnAlaIhrPheThrAspTrpTyrAlHisAspMet	643
Qy	1798	TTAAGGAGTACTTAATGATGGAATGACAAACTTAACCTTGATTAAT-----AATGCA	1851
Dd	644	LeuIleAspSerIleIlySTrpArgAsnGlnHisGlyLysCysIleAsnIlyAspAsnGly	663
Qy	1852	ACCACGCAATTTGTATGTATGAATGATTAACAGAAATTTGCTTAATGTTTGCACAGATGGGTTAA	1911

[illegible]

```

QY 3781 -----TGTAAACAGTCGAT 3795
Db 1336 GluProAlaAlaIleProThrThrThrProGlyValThrProAlaCysGluIleVallys 1355
QY 3796 GATATACCTTAA-----GAAACGATGGAAGAACAGTACAGATTTCATCCAAA 3849
Db 1356 AspleuHeGluIulysProIylsasnThrPheLys-----GluAlaCysThrGlnLys 1372
QY 3850 ----AAGAAATGATGATCCGATTCGCAATGC----- 3882
Db 1373 TyrGlyGlyLysasnSerArgLeuGlyTrpLysCysIleLeuProThrSerGlyGlyAspLys 1392
QY 3883 -----GGAATATATAATTGATGAGACCCCTGCT----- 3912
Db 1393 AlaAlaThrArgGlySerGlyAspThrThrLysGlnAsnAspSerGluGlySerGluGly 1412
QY 3912 ----- 3912
Db 1413 GluGlySerHisGlnArgAlaLysArgHisThrSerAspAlaSerGlyGlyLysSerAla 1432
QY 3913 -----GTGTATATGCCCTTAGAGACAAAGTTATGC 3945
Db 1433 LysSerGlyGlyIleProThrGlyGlySerIleCysIleLeuProArgArgArgLysLeuTyr 1452
QY 3946 GTACATTCTTGGCAATGATTAATGAATTAATAAATTAATCACAACAGTT----- 3996
Db 1453 ValGlyGlyLeuThrLysTrpAlaGluIleGlnSerSerInSerGlnAlaLeuSerGly 1472
QY 3997 -----AATTTAAAGAAAGCTTCATC 4017
Db 1473 GlnThrThrProAlaGlyThrProSerGlnAlaGlnAspProLeuAlaAlaPheVal 1492
QY 4018 AATTCGACAGCAGACAGAAACATTTCTCATGGTATTATTATAA----- 4062
Db 1493 GluSerAlaAlaValGluThrPhePheLeuTrpAspArgTyrLysLysLeuAsnAlaPro 1512
QY 4063 ----AGTAAGATGTGGAAGAAATGAATC----- 4089
Db 1513 GlnSerGlySerSerLeuGlyGlyAlaProLeuGlnLeuLeuAsnGlyAlaIleGly 1532
QY 4090 -----GATAAAGATTAAGAAAGCGCAAAATTCCTCCGCAATTTTGA 4134
Db 1533 SerGluGluThrProGluThrSerLeuLysSerGlyThrIleProProAspPheLeuArg 1552
QY 4135 TCCATGTTCTACACATTGGAGATTATAGAAATTTTATTATTTGACA-----GATATA 4188
Db 1553 LeuMetPheThrLeuGlnLysAspTyrArgAspIleCysValGlyValLysGlnAspVal 1572
QY 4189 TCAAAA-----GGTCATGTGAGGAGGAGTAAA-----CTA 4218
Db 1573 IleLysAlaLeuGlnLysAlaSerGlyAspAsnLysSerSerLysAsnProMetGlnGluIle 1592
QY 4219 AAGAGCAATATAGATCTCTTTCAAAAATGAGACCAAAATCTCT----- 4266
Db 1593 SerSerLysIleGluGluIleLeuLysAsnGlyThrProProThrProThrProValThr 1612
QY 4267 ----AATGAAAAACACGCCAAGAAATGCTGACAGAACTACTGATGATA 4314
Db 1613 HisSerProSerSerGlyThrThrProSerSerTrpTrpLysThrAsnGlnHisIle 1632
QY 4315 TGGGAAGCTATGCTATGCTAGTACAGTA-----AAATTTGGCGCAAAAAA----- 4359
Db 1633 TrpLysGlyMetIleCysAlaLeuThrThrLysGlnSerIleGluLysIleGlnGln 1652
QY 4360 ----GATGATTTTACGAAACTAGCGTTACAAACGCAATTTAGTGC--- 4407
Db 1653 ValLysAlaThrAspAsnThrAspLeuPheGluLysLeuLysAspLysTyrSerAspTyr 1672
QY 4407 ----- 4407
Db 1673 AspLysValGluIleLysGluGlnLysGluThrGlnAlaLysGlyProGlnAspGlyLeu 1692
QY 4408 ----AAAGCACCACTTTGGAGCAATTTGCCAAAGACCCCGATTTTACGATGGCTA 4461

```

```

Db 1693 ThrProGlnThrThrLeuLeuSerAsnPheValLysArgProProIylPheArgTyrLeu 1712
QY 4462 ACCGAATGTCAGACAGCACTATTGCTATACAGCAAAATAATTTGAAGATGGCAGAA 4521
Db 1713 GluGluTrpGlyGlnAsnPheCysLysThrArgLysArgMetLeuLysAspIleIleTyr 1732
QY 4522 AATGTAACTCAATATGACCAATTTGAAGTGTGAT-----ACAGAAATGT 4563
Db 1733 GluCysArgAsnSerAspAsnProGlnHisAspTyrCysSerGlyAspGlyTyrHisCys 1752
QY 4564 AATTAAGAA-----TGCAGAGCTACGTTAAA 4590
Db 1753 AsnAsnGluLeuLysHisAsnAsnMetPheHisGlyIleAspCysArgAspCysTyrLys 1772
QY 4591 TATATGAAAAAAGAGTGGATTCACAAAGATTAATTAATTAACAAAGATGAACGGCAG 4650
Db 1773 GluCysArgLysTyrLysLysTrpVal-----HisLysLysPheAspGluPheHisAsnGln 1791
QY 4651 AAAAAAGATTGATACACACACATTTGTAATGCTTACAGAC---TATACTGCA--- 4704
Db 1792 LysAsnLysTyrGlyGluGlnHisGluLysLeuThrAsnGlyAspAsnTyrSerGlyGly 1811
QY 4705 ----ACGAATGCAACAGATTACTTGAACAGAAATTTACTGCTACTTGTGCTGAT 4755
Db 1812 GlyAspAsnThrAsnPheCysGlnGlnIleLysGluLysLysThrAla----- 1827
QY 4756 AAGCCTGGAAGTGCTGTGTGTAACAAAGAAATATACAAATTTGTAAGAAACAGGCTTAC 4815
Db 1828 -----GluAspPhe 1830
QY 4816 TATGATGCCGACAAACATTTGGGTGCACAAATTTATTAAGAAATGACGCAATATATCT 4875
Db 1831 LeuLysAlaLeuThrArgHisCys-----LysAsnSerGlnAspAsnThrAspLysSer 1847
QY 4876 AACATTTGAGTAAAGTAAAG----- 4896
Db 1848 GluGluAspGluLysAsnLysIleAsnPheAspLysProGluAsnThrPheAsnProSer 1867
QY 4897 ----TGCAA-----GGATTTGTAAGAGCAACACAGCGCTATT 4935
Db 1868 ThrTyrCysLysAlaCysProIleTyrGlyValThr-----CysAsnArgGlyThrArg 1885
QY 4936 AAGTGCAGAAACAAAGGCTCTATATACATCAATCACTTGAAA---GAATTCAGTGA--- 4989
Db 1886 GlyCysArgProLysIleIleAsnGluLysAsnAsnIleGluGlnGlnThrAspIle 1905
QY 4990 GATGTGCTTTTCTCTCTGCTGCTACGTATATGTTTCATGATGATGCAATTAAT 5049
Db 1906 AsnIleLeuLysAsp-----AspGlyAlaThr 1914
QY 5050 ACAGATCCAGAAATTAAGATGAATAATGGGTGCGAAAGAAATGATGAGAAAGCGCGCA 5109
Db 1915 AsnAsp-----ThrAspAsnGluLeuHisLulysCysLysGlu----- 1927
QY 5110 ACGGAAGGTACAATTTGGCTCAATTACTACAAAGAAAAAAGAAAAAGAAAAATATAA 5169
Db 1928 ----TyrGlyLeuTyrThrAsnLeuLysGlnGluTrpLysCysGln 1942
QY 5170 ACGTCGATGCGACAAA---TATTCATTAGAGGTCCCGCTGTACTGCTATCAATAT 5226
Db 1943 ----HisLysAsnLysSerTyrLysCysGluLeuGlnLysProLeuAsnSer 1958
QY 5227 AGTTTATGATTTAAGA-----GATATA 5250
Db 1959 GluTyrTyrAspGluArgIleProPheLysIleLeuPheGluArgTrpIleIleAspPhe 1978
QY 5251 ATTCTAGTATTGATAATTGGAAGATCAAAACAAAGACCGAGGAAATTTGAAGAAA 5310
Db 1979 Ile-----GlnTyrTyrAsnLysSerLysGluArgGlyIleThrArg 1991
QY 5311 ATATTTAACAAATAATGAACATCACTGTCGCAAGAGAGTGAATGACTACAGGAAATCCC 5370

```

QY 97 AAAATGCAAGAAATGTTTTGAACGGTTATGCCAAAATATAAGACATTCATCCAAAA --- 153
 ::||| :||||| :||||| :||||| :||||| :|||||
Db 12 GUSerValLysSmetPheaspArgILlegIyGIAsPvalTyrGlugInVallysser 31
 ::||| :||||| :||||| :||||| :||||| :|||||
QY 154 TATGC AAAAAGACAATGTGGATTCGTGAAAGGGGATTTCACNAAGCAGAATTCGTGGT 213
 ::||| :||||| :||||| :||||| :||||| :|||||
Db 32 GIurThrValasnyTyrValSerGIureuGIugLIyLysLenSerLeuaAlaProIIeleucGIy 51

```

Oy 214 GGTCTTCTAGCCGAGTAATTAAGCTAAATTATTTATATCAATATCCATGTAAATTAGAT 273
    |||
Db 52 ValGlsuSerGlySerThrHisnglu -----ThrCysAsnLeuVal 64
Oy 274 CATAGGAACATCTCTAATTATTCAGTATGATGTCGAAATTTCAGACATCTCTGCACGT 333
    |||
Db 65 GluAspArgTyrTyrAsnLysProValTyrGluAsnSerAsn ---ArgTyrProCysLysAsn 83
Oy 334 -----AGAGACAAAACCGATTGTATGAGAAAGATGAAAGAAATGTGAAATGTGAAT 381
    |||
Db 84 LeuLysGlyTLeThrAsnGluLutArgPheSerAspThrLeuGluGlyGlnCysThrAsn 103
Oy 382 AAAATA-----CGTATATATTAAGAAAAAAT-----GATGCTATAGCCTGTGC 426
    |||
Db 104 LysLysTLeLysGlyAsnGluTyrTyrSerThrTyrSerGluLysAspCysGluTyrAlaCysAla 123
Oy 427 CCACCTAGAACAGCAGCATATGTGTGATAAACACTGGAAAGCTCTAAATATGATATAATACC 486
    |||
Db 124 ProTyrArgArgLeuHisLeuLysSerHisAsnLeuGlnSerLeuAsp -----ThrThr 141
Oy 487 CAATAATATCAAGATTATTTATGGAAATGTACTACTTACAGCAAAATATTCAGAGGTAAATCA 546
    |||
Db 142 SerMetThrHisLysLeuLeuLeuGluValAlaCysMetAlaAlaLysTyrGluGluGluAsnSer 161
Oy 547 ATTGTATTAATATCATCSA---CATAAAGCAAT-----CTACAGCGCTGT 588
    |||
Db 162 IleAspThrHisTyrProGlnHisGluArgHisAsnGluAsnSerProSerGlnIleCys 181
Oy 589 ACGGCGCTGGACGAGAAAGTTTGGCAGATATAGATATTTGTAAGAGAAATACATATGTT 648
    |||
Db 182 ThrMetLeuAlaAlaArgSerPheAlaAspIleAspIleValAlaArgLysLysAspLeuPhe 201
Oy 649 AAACCAAAAT-----GTCCATGACAAATATAGAACAGCGGTCCGAGAGGT 693
    |||
Db 202 TyrGluLysSerTyrGluLysGluLysArgAspGluLeuGlnIleuThrHisLeuTyrThrIle 221
Oy 694 TTCAGAAATATACATGATGAGCATGCAAGATGAA-----GTAAAAATGATTACAATCCT 747
    |||
Db 222 PheGluLysIleHisGluLysLysLeuLysAspLysGluGluYalaGluLutArgTyrGlySer 241
Oy 748 GATGAGTATGGAATTTATTTAATTTTAAGAGAAAGCATGGTGAAGTGAATGAATAATAA 807
    |||
Db 242 Asp---ThrThrAsnTyrTyrGlnLeuAlaGluLysPrtTyrTyrAlaAsnArgAlaThr 260
Oy 808 GTATGGGAAGCATATAACATGATGATGCATCATATATAATCTGGATTTTATTCGAATCA-- 864
    |||
Db 261 ValTyrGluAlaIleThrCysAspVal---HisGlySerAspTyrPheArgGlnIleThrCys 279
Oy 865 -----GAAGTAAATACACCATTAATTTTCAAAATCCTAATATGC----- 900
    |||
Db 280 GluAspLysGluThrThrAlaIleThrArgValLysAspLysCysArgCysLysAspGluAsn 299
Oy 901 GCCCATTAACAAAGA-----AAGGTCCTACCAATTAAGATTATGATCTCCCTCAA 948
    |||
Db 300 GluLysLysProGlySerAlaSerAlaAspGlnValProThrTyrPheAspTyrValProtein 319
Oy 949 TATTACGTTGGTTCGACGAAATGGGAGAAAGATTGGCCGAAAAAGAAATATTAAATG 1008
    |||
Db 320 TyrLeuAlaArgTyrPheGluGluLutPrrAlaGluAspPheCysArgLysLysLysLysLeu 339
Oy 1009 AAAAAGCTCAAGACCTCTGTCTGTAATGCAAAAGAACGCTTATATTGTAGTCATATGGA 1068
    |||
Db 340 GluLysLeuGluGlnGlnCysArgAspArgTyrGlnAlaHisLeuTyrCysSerGlyAsnGly 359
Oy 1069 CATGATGTGTCAGCAACATATTGTGGAAAAAGGATTTTGCATTGTGATATATAGTACT 1128
    |||
Db 360 TyrAspCysThrTyrThrIleTyrTyrLysGlyLysLeuValIleGluGlnHisCysThr 379
Oy 1129 GACTGTTCGACTAAATGCAAAAGTTTGTGAAGTTGGGTTAAGGAAATCAACAAAGACATTT 1188
    |||
Db 380 AsnCysSerValIleProCysArgLeuTyrGluLysSerTrpIleAspAsnGlnGluLysLeuPhe 399
Oy 1189 AAAAAACAATAAATAATATGAAAAAGAAATACAAATCATATTTTATGC----- 1236

```

Db	400	LeuylsclnlnysglnlnysTyrGlutrlGluIleSerAsnSerclYserCysglYclYser	419
QY	1237	-----AACGTAACAATTTGTCATATATTAAATAGTGA-----	1272
Db	420	GlyglYallYsGlYargAsnArGlyLysArGlyAlaGlyValGluIuThAlaerAsn	439
QY	1273	-----TATTATTAACAATTTTATGAAAACTTAAGGAAGGCATATGCAACTAT	1323
Db	440	TyrAspclYtrGlYlulYslYsPheIrrYlYsGluIuLeuYlSglUserGluTrYclYlYsVal	459
QY	1324	GACACTTTTAAATTTACTAAATGAAGGAAGATATTGTGAAGAGAGATTTCCAGAGAA	1383
Db	460	AspAspheleuYlSleuLeuAsnAnSnglUasPvalCysLysLysIleLysAspclulYs	479
QY	1384	AAGGATTTACTTTTACTTAACAGTGTAT-----GACAAAGGATATTTTAT	1431
Db	480	GluYlSleAspheIrrYlYsProAlaAsPlyAsnSerAsnAnSnglUglYthrPheYtr	499
QY	1432	CGTTCAAAATATGCCAAGTGTGCTCCGACCTGGCGTCAATATGATGATATAAAATC	1491
Db	500	HisIrrclulYrrYslYsProCysProAsPcYsGlyValYlArGlyAspAnSlnIrrP	519
QY	1492	ACACACAAATACAGATATGATGCTGTGAACGTGAATATGAAGACTATTAACCTGCATGG	1551
Db	520	LysAspLysYlrrYsPclYlYsCysThrArG-----GlyLysLeuYlrrGluProAlaser	537
QY	1552	GGTGTGAAGCTACTATATACACTGTCTTATAGTGATGAACAAGATGATATACA	1611
Db	538	GlylAclnglYhrProIleuYlSleleuYlSserGlyGluYlSglulYlSglulrrGlu	557
QY	1612	CAAAATATCAAAATTTTGTATAC-----AGTCACACTAAT-----	1647
Db	558	ThrLysLeuYlSAlaPheCysAspGlnThrAsnGlyAsPThrThrAsnSerValAlaArG	577
QY	1648	-----TACAAGATTAATAATTAACAATA-----TGG	1674
Db	578	GlylYclYlYlAlaAspclYserGlySerLysSerAsnSerLysGluLeuYlrrGluIuTrP	597
QY	1675	GAATGCTATTATAG-----	1689
Db	598	LyscysYlrrAsnGluValGlnLysValLysAspAsPlyAsnGlyGluIuIuIuAspLu	617
QY	1690	-----GATGAATAATTAATAGA-----TGTAAACTGGA-----	1719
Db	618	AspclnGluIuAsPvalAsPlyValLysLysAlaGlylYleuCysIleleuGluIuAsnLys	637
QY	1720	-----CAAAATCTGAATACATATGATATGATATCCCAATATATATCATTCAT	1767
Db	638	LysHISglUserArGysAnSnsSerSerAsnGluProGluIuPheGlnLysThrPheHIS	657
QY	1768	AATTTTGTGAATTAATGCTTACATATTTTATTAAGGATACATTTTAAGTGATACAA	1837
Db	658	AspheheYlrrPheIrrPleGlyArGpHeleuAsnAspSerMetclYlrrArGlyLys	677
QY	1828	CTTAAAACTTGTAATAATAT--ACACACAGCATTTGATTGATGAATGTAAACGAAT	1884
Db	678	ValAsnSerCysIleAsnAnSnpProLysArGlyLysCysArGAsnGluIuclYlSAspAsp	697
QY	1885	TGCTTATGTTTGCACAGATGGCTTAACAACAAAAAGAAAGATGGAATrATrTAAAGAA	1944
Db	698	CysglYcysPheLysglUtrPleGlyLysLysLysGluIuTrPclUAsnIleLysLys	717
QY	1945	CTGTTCACAAAAAAGAAATATACAGCAATGATATTATAGTAATTAATATCTTTT	2004
Db	718	HisPhe-----	719
QY	2005	GAACGTTATTTTAAAAAGTTATGATTAACCTTGACAAAGATGAACAAATGCAAGAA	2064
Db	720	-----LysThrGlnGlu	723
QY	2065	CTTATGGAATATTAATAAAGAAAAAATATGACTTTTCCATTTTGGAATAATATAGGAC	2124

Db	724	AlaPelysAen---LysatrgUlaunSercJylleasmpetHesercJyleuNetasp	742
Qy	2125	TATTAGGAATGCATATAGAACTCTTGTTGATCACTTAAAGAAACTGCCAGATATGT	2184
Db	743	SerAlaaspValValleugJuleuAlaleu-----	752
Qy	2185	AAAGCACAATATACAAACGAAGCATGTGAAATCCCATATATGCAACAAACCCGTGT	2244
Db	752	-----	752
Qy	2245	GTTAAACCTCGTGGAGGCACGCAACCCACTTAAATATATAAGAAATAGCACAAATCTT	2304
Db	753	-----	758
Qy	2305	AAAGGAGTGCATACGAGGAACGCAAAATCGTGTCTTCATTAATTGAAGAAAGCA	2364
Db	758	-----	758
Qy	2365	CACGAAGTATATATTAACCTGGGGGTAGGAGAAAGACTTCAGAGCAATTTATGTAGA	2424
Db	759	-----	763
Qy	2425	ATATGATAAACATTTCTATCGTAATCTGTGTTTCAAAATGACCATGTGTCGAAA	2484
Db	763	-----	763
Qy	2485	GGCAGAGTGTAGGTATACAAACAGATTTGTGTAGAACTGAATGAGMAGTGCATCG	2544
Db	764	GlyTyrgIcYasp-----VallysgJuleuLysGlyllyleuSgluleuLeuaspGlu	780
Qy	2545	GAACACATGCGTAAAGATCACGAAGATGTTATTATGCTCTTCAGAAAGACACATATATGT	2604
Db	781	GluLysLysLysLysGlnalagJugJualValVal-----	792
Qy	2605	ACATCAATTTGGAACATTTACAAACGAGATGATCACCCACTTAATGTATATTTGTGT	2664
Db	793	-----	794
Qy	2665	GATTAGTATAATATTCCTTTTGGGGAGATGTTCTTCTATCAGCAAAATATGACAAAC	2724
Db	794	-----	794
Qy	2725	AAGATATAGCAATGTATATAAGAAAGAAATTAACCTTAAGGGCCCCAAAGAACTATGAC	2784
Db	795	-----	797
Qy	2785	CCAAACACCGACACACTATCTGCGAGTATACGTTTACAGTTTGCACATATATAGTGT	2844
Db	798	AsngJnLysLysTrnTrpIleasPlyLsleuJGlnHlsGlu-----	811
Qy	2845	ATAAATCGAGAGAGATCTCTGGGAAGAAAGAGCGATGATGTAAGCTGCAGAGCAT	2904
Db	811	-----	811
Qy	2905	TTGGAAACTGTTTTTGGTAAATATACATTAAGTCACTCAAGGCAAAAGAAATGATAATAT	2964
Db	811	-----	811
Qy	2965	AATGATGATGCCCCCAATATTTTAAATTTAGGAAAAATTTGGGAGACATATAGAGCC	3024
Db	812	GlyaspaspAlaasnaasCysLeuLysTrpHls-----	822
Qy	3025	AAAGTATGGAAGCATGAATGTGATATATAAATTTTGAAGGATTAATCGGGACACAA	3084
Db	822	-----	822
Qy	3085	TCAACACAAAGTACTTATTCGCGATATAGTATCATACACATTTGGATGATTTATCCCA	3144
Db	822	-----	822
Qy	3145	CAAAATTTAAGATGATGATGACCGAAATGGCGCAATGGTACTGCAAGGTGCAGAAAAAGAG	3204
Db	822	-----	822

QY 3205 TATGATAAGTGAAGAGAGTGTAGAGATGTAAGATTAAGAT-----AATGCT 3255
 Db 823 -----LysGluLysCysGluLugluThrGlnProLysProProGlyAlaGlyGly 838
 QY 3256 CAAGCTGTACGAAGAGAGTGTACAGAGTGTACAGAGTGTACAGAGAGTGTATGAA 3315
 Db 839 ProGlyAlaProSerGluThrGlyLugluThrThr----- 850
 QY 3316 TATAATGATATATAGATATGTAAGAAACATGATATATATCATGATTAATACAA 3375
 Db 851 -----LeuGluAspGluGluGlu 856
 QY 3376 GAATTACATGACAGACACAAATCTGTGTATGTAATAGTGTATGAGCTTCCAGTACT 3435
 Db 857 GluLugluAspGluGlu-----GluAspAlaGlyAspGluVal----- 868
 QY 3436 GCCAAAATCATATAGACAGAGATGTATGAAATTTTTCGGAATTTATACCAACAAAT 3495
 Db 868 ----- 868
 QY 3496 GGTGGCAAAAGTAAAGTGTACTAGTATGATGAAAGTGTCTGATGCTACTTAACACC 3555
 Db 869 -----GluGluGlyLugluThrValAspThr----- 876
 QY 3556 ACGTATGAAATGTGTGAGCATATCTCCATGATACAGAAATTTGTGATGTCTAGTCA 3615
 Db 876 ----- 876
 QY 3616 CAAATGAGTTTGTGATGAAAAAGTGTAGATGTAAGATACGAAATATGCTTTAGA 3675
 Db 877 -----ThrGluGlyAspGluThrGluThrValGluGluInPro 888
 QY 3676 GATAAACACAGACCATGATGTGCGTGTGTGTGTAAGAGTGAGTACGAAACCGACAGG 3735
 Db 889 ValLysAspThrAspArgGluGly----- 896
 QY 3736 GTACAGATTAAGAAAAAGGCGAAGAAAGATACGAA-----TGTAA 3786
 Db 897 ---GluGluGluGluLugluLysLysAlaThrAspThrThrThrSerLeuAspValCysAsp 915
 QY 3787 ACGTGAATGATACTTAAAGAAAAAGATGGAAGAAACAGTACGATGCTATCCA 3846
 Db 916 ThrValLysAsnAlaLeuThrAsnAsp-----AsnLeuThrAspAlaCysLysLeu 933
 QY 3847 AAAAAGATATGTAATGCA-----TATCCGATGTGCATGC-----GGAAT 3888
 Db 934 LysTyrGlyProGlyLysGluThrGlnProAsnThrLysCysValSerSerGlyGlu 953
 QY 3889 ATTAATTTAGTG-----GAAAGACCTGCT 3912
 Db 954 LysSerValAlaThrAlaGlySerSerGlyAlaThrGlyLysSerGlyAspLysGlyAla 973
 QY 3913 GTGTGATGCCCTAGAGACAAAGTATCGTACATTTCTTGCGAAATGATATGAA 3972
 Db 974 IleCysValProProArgThrArgThrLeuThrValGlyGlyLeuThrLys----- 990
 QY 3973 ATTAATAAATTAACAATCAACAATTAATTAAGAAGCTTCAATCAATTCGACAGCA 4032
 Db 991 -----LeuThrSerAlaGlyThr 996
 QY 4033 GAAACATCTTCTCATGTATATTAATAAGATGAGTGTGAGGAAGATGAACTGCAT 4092
 Db 997 SerSer-----GluSerProGlnGly-----GlySerGluSerSer 1008
 QY 4093 AAAGATTAAGAAAGAGCAAAATCTCCGATTTTGAATCATGATTTCTACACATTT 4152
 Db 1009 Arg----- 1009
 QY 4153 GGAGATATAGATTTTATTTGGAACAGATATATCAAAAGTACGATGAGGAGAGT 4212
 Db 1010 -----AlaSerAspValSerGlnGlyAsnGly----- 1018

QY 4213 AAACATAAGAGCAATAGATTCTCTTTTCAAAAAATGTGACCAAAAAATCTCTAATGGA 4272
 Db 1018 ----- 1018
 QY 4273 AAACACGCCAAGATGTGACAGAAACATGATGATATGGAAGCTATGCTATGCT 4332
 Db 1018 ----- 1018
 QY 4333 GCACCTGTAATAATTTGGCGCAAAAAAGATGATTTTACGAAAACCTACGTTACAAAC 4392
 Db 1018 ----- 1018
 QY 4393 GTCAATTTAGTCAAAAAGCAACCACTTGTGAGGAATTTGCCAAAGCCAGTTTGA 4452
 Db 1019 -----GlyAspAspLeuThrThrGluSerLeuThrLys----- 1030
 QY 4453 CGATGGCTAACCGAATGTAAGCAGCAGCTATGCTATACACAGCAAAAAATTTGAAGAT 4512
 Db 1031 ---TrpPheIleGlu----- 1034
 QY 4513 GTGCAGGAAAAATGTATAGTCAATGACCAATTTGAAGTGTATACAAATGTAATAGAA 4572
 Db 1035 -----ThrAlaAla 1037
 QY 4573 TCGAGAGACTACGTTAAATATGAAAAAAGAGTGTATCCAAAGATTAATAT 4632
 Db 1038 IleGluThrPhePheLeuThrPheLysArgThrLysGluThrGluAlaGluLysLys--- 1056
 QY 4633 TACAAAGATGAACGCGACAAAAAAGATTCATGACAAACATTTGTTAGTGTACA 4692
 Db 1056 ----- 1056
 QY 4693 GACTATACTGAAACGATGCAACAGATTAATGACAGGAATTTACTGCTAGTTGTGCT 4752
 Db 1056 ----- 1056
 QY 4753 GATTAAGCTGGAAGTCCCTGTGTGACAAAGAAATATACATTTGTAGAAAAACAGCT 4812
 Db 1057 -----AlaGluLeuGlnLugluAsnGlyLeuLeuLeu----- 1066
 QY 4813 TACTATGATGCCAACAATTTGTGTGTCACAAAATTTATTTGAAATGACGACAAATAT 4872
 Db 1066 ----- 1066
 QY 4873 ACTAATTTGAGTAAGATTAAGTCAAGATTAAGTAAAGAGGCAACACAGGTGCT 4932
 Db 1067 -----GlyThrGlyAla 1070
 QY 4933 ATTAAGTGGCAAAACAAAGTCTTAATTAATTAATTAATTAAGATGACTGAAGAT 4992
 Db 1071 Ser-----LeuAsnLeuGlyGlyAspAspSerAsnProGlnThrGlnLeuGlnLysSer 1088
 QY 4993 GTGCTTTTCTTCTCGTGCAGCTAGTATAGTTTTCATGATGATGATGCAATTAACA 5052
 Db 1089 GlyThrIleProLeuAspPheLeuArgLeuThrLeu-----GlyAspTyrArg 1107
 QY 5053 GATCCAGAGTATAA-----GATGAATAATGGTTGCGAAAAAGATGATGAGAA 5100
 Db 1108 AspIleLeuValArgGlyValAlaAspAspLysAsnGly----- 1120
 QY 5101 GTGCGGCAACGAAAGGTCATTTGGTCACTACTACAAAGAAAAAGAAAAAGAG 5160
 Db 1121 -----GlyAsnAsn----- 1123
 QY 5161 AAATAAAGAGTCGATGCGCAAAATATTTATGAGGTCCCGCTGTAGTGTATG 5220
 Db 1123 ----- 1123
 QY 5221 AAATATGTTTTATGATTTAAGAGATATATCTAGATTTGATTAATTTGGAAGATGA 5280
 Db 1124 -----IleIleLeuAsnAlaSerGlyLysLysAspGlu 1134
 QY 5281 AAACAAAAGACCGAG-----GAAATTTGAAGAAAAATATTTAACAATAATGAGACA 5331


```

|||||  |||  |||  ::::::::::|  ::::|
Db 1135 LysGlnMetGluLysIleGlnGluLysIleGluGlnIleLeuProThrSerGlyAsn 1154
QY 5332 TCAAGTGGCAAGAGAGATGATGACTACTACAGAAATCCCGTAGTACTGGCGCAAAATTT 5391
Db 1155 LysGlnThrArgGlyProGlnAsnSerValAsnAsp-----ArgGlnSer 1169
QY 5392 TTCTGGACGAAATTAAGAAATGTGTGGACGCAATGATATCGGG-----TACAA 5445
Db 1170 LeuTrpAspArgIleAlaGlnHisValTrpHisIleGlyMetValCysAlaLeuThrTyrLys 1189
QY 5446 CGTGTGAGATGATGAAATAGTGGAAATAGTCCAAAGAAAGTGTGAAGATCTAAAAA 5505
Db 1190 -----AspAspAspAsnGly----- 1194
QY 5506 TGTGCTTCTGACCTTCAGATGATGATTATCTATGGGCAAAATCCGATGAAGTACT 5565
Db 1194 ----- 1194
QY 5566 GCGTATCATGTTCTTCGATGCTTTGCCGAAATGGGGTGAAGATTTTGCACAAATAAAGAA 5625
Db 1194 ----- 1194
QY 5626 AAGGAATTGAGAAATGTGTAGGGCGGTGTAATGATTATTACTTGTGTAATGAAGAT 5685
Db 1194 ----- 1194
QY 5686 AAAAGAAAGAAATGTACAGATGCGGTGACACAAATATTAATAATTTATAGTAGTGAAA 5745
Db 1194 ----- 1194
QY 5746 CCACAGTATGAAAAACAATCAAAAAATATGTGAGATTAAGACAAAAATATATTCGAG 5805
Db 1194 ----- 1194
QY 5806 CATCTGTGGCAAAAGATGAGAGAGACGCTCGCAATATTTAGCAAAACAATTAATAAAA 5865
Db 1194 ----- 1194
QY 5866 ATTGTGAAATATAAAGTGAAGATTTGAAATATATAGTATGAAGAAGATGTGCCACACAG 5925
Db 1195 -----LeuLysGlyVal----- 1198
QY 5926 CGATTAACTGATGTAATAGTCAAAATATATGCCGCATCATTAAGCATGAACCAAAAGAA 5985
Db 1198 ----- 1198
QY 5986 GTTGAAGGAAAGTGTATATGTCAAGTGCACAGAGTCCACACGCTGTCAGAAAGGAAACA 6045
Db 1198 ----- 1198
QY 6046 CCGTCACACAGGGTATCATGATATCAAAAGCGACGCAATCGAAAAAGAAAGCGAAAAACA 6105
Db 1198 ----- 1198
QY 6106 GCCCGGCTTACAAACACCGCAAAAAAGTGAATAATCTAACACAGAAATGCCAGCAACA 6165
Db 1199 -----ValLysLysProGlnLysIleGlnAsn----- 1207
QY 6166 ACACGAACCCGACGAGACACAAACAAACAGAAAGAACATCAACAGACAAACAACA 6225
Db 1207 ----- 1207
QY 6226 GAATCTGACGTGGCACAATGTAAAGGCCATTCCTTTCGAATAAACAGATAGCAGGGCT 6285
Db 1207 ----- 1207
QY 6286 GGAATAGAGGTTGTATCCAAAAAGATATGACAAATATCTTAATGGGTTGTATTGTA 6345
Db 1207 ----- 1207
QY 6346 GGTAAAGTCTAAAGAAATGAAATATGCAATATGATGCTCTAGAGAAAAAATTTATGT 6405

```

```

Db 1208 -----ProGluLys----- 1210
QY 6406 ATAAATATATACATATTTAAATTAAGAACTGAAATAAGCGTCGAATGATATATAA 6465
Db 1210 ----- 1210
QY 6466 GAGCCTTTATTAATGTGCAGCATATGAAACCTCAATTTTGTGTTAAATATATATTT 6535
Db 1211 -----LeuTrp----- 1212
QY 6526 GAAATTCCTGACGAGAAAAATGAATTCGAAATGGAACAATTCACAGATGAATTTAAAGA 6585
Db 1212 ----- 1212
QY 6586 ATAAATGATATATACATATATGATTAATTAAGATATGTTTTTGGAACTGATATTTCTAAT 6645
Db 1212 ----- 1212
QY 6646 GATAAAAAATTAATTAAGTATGTAACAAATAGTGTAAACCAATTCATGAAATTAATAG 6705
Db 1213 -----AsnGluThrThrLys----- 1217
QY 6706 AAAAAACAGATTAATAAAGATGAAAGAAATTAAGTAATAATTTTGGGAAAAAATAAA 6765
Db 1218 LysProLysAspGluLysTyrGlnTyrGlnThrAlaLysLeu----- 1231
QY 6766 AAATTTATTTGGAGAAATGATATATGATTAATTAATCATCTACACAGCAAGAAACGAA 6825
Db 1232 -----GluAspGluSer----- 1235
QY 6826 AAGAAAAAATTAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 6885
Db 1236 GlyLysLysArgProAspSerSerAlaSerGly-----ThrLysLeuThr----- 1250
QY 6886 GAAAGTTTGTAAAAAGCCCCCAATTTTGTGAGAGGTTTCACAGAAATGAGGCAAGAAATTT 6945
Db 1251 -----AspHeiLeuLysArgProPoriLysPheArgTyrLeuGlnGluLysPhe 1269
QY 6946 TGTATATAGAGGAAGCAAGTGTGTAAATTTGAGCGCGGCTGTAAAGATATAGTGT 7005
Db 1270 CysLysLysArgThrGlnMetLeuGlyLysIleLysGluAspCys-----TyrLysAsn 1287
QY 7006 AATGTATGATATGACGTTAGACACAAAGATGTCCAGAG----- 7044
Db 1288 GlyLysArgCysSerGlyAspGlyLeuLysCysAsnGluIleValIleAspLysGluLys 1307
QY 7045 -----GCGTGTATACATATCAAAATTTTATTAAGAGCG 7080
Db 1308 IlePheGlyAspLeuLysCysProThrCysAlaArgHisCysArgPheTyrLysLysTrp 1327
QY 7081 AAAATGAAATATGAAGACAAAGAAAGTTCAAAAAGATTAAGATGCGCAAAAGTAT 7140
Db 1327 ----- 1327
QY 7141 AAGATTAATCTTCTACTGAAAGACATAGAGAAGCAACATGTGCTCATGAATATTTA 7200
Db 1328 -----IleAsnThrLysArgAsp----- 1333
QY 7201 AACATGAAATTTAAAGAAATTTGTGGCAATTAAGGATTTGTCTGTATGCAAAAAACCTCT 7260
Db 1333 ----- 1333
QY 7261 TCACAACTACCAAAAAACAAACAAACAAATCAACATCATCCGATGCTAATGATATGCCAGAA 7320
Db 1333 ----- 1333
QY 7321 TCAGTGAATATGTCTTCTGAAAGATTTAACAAGTGTAGTCTCTGAACCTTTCAAAAAAG 7380
Db 1334 -----GluPheAsnLys----- 1337
QY 7381 GGATCTATGATTCATACAAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 7440
Db 1337 ----- 1337

```

QY 7441 AAAGCAGCATATATTATCTAAGAGCAAGAAATATATGATATACCTTTGAAGGAA 7500
Db 1338 -----GlnSerAsnAlaIleTyrSerGluInlYsLys 1347
QY 7501 AAATTTATACCTATGAGTCTACAAAGAAAGAAAGAAAGTAAATATGTCGATATAT 7560
Db 1348 LysTyr-----GluGluGluAsnAspSerAlaGluInlYsAsnAsn 1360
QY 7561 AATCTTGGCAGCTAAGAAACCTATGACCTGATTAATATATGAGAGA-----AGA 7614
Db 1361 GlYAlaCySgLYThrLeuLysAspAlaIaGlu-----PheLeuAsnThrLeuLysAsn 1379
QY 7615 AACCCCTGGGAAATATAGAGAAAGAAATGCTTTAAAGTAGATTTAAGATGAAATGTTAC 7674
Db 1380 GlyProCySAsnGluSerGluInlYsAsnLysAlaGluAspGluIleAspPheLys 1399
QY 7675 AAAAATTCAAAGTCTATCAGAGAGAAAAAGATGATGTGATGCTCCAGAGAGCAACAT 7734
Db 1400 LysProAspPThrPheLysAspAlaAspAsn-----CysLysPro----- 1413
QY 7735 ATGTCTTAAGAGATTTTATGATGAATTAATTAATGAAAGACTTAAGATGATTAATATCTC 7794
Db 1414 -----CysSerGluPheLysIleLys----- 1420
QY 7795 CTAAATATGTTCTGCGCACTGCAAGAAATAGAGAAATAGCATATATTAATTAATCTCAC 7854
Db 1421 ----- 7854
QY 7855 TCAGAGAACGGGTCCGCAATGCAATATATGAT-----ACTATGAATATATGTTGCTGCT 7911
Db 1426 CysSerSerGlyGlyAsnThrGluGlyLysCysAspGlyLysThrThrIleAlaIaThr 1445
QY 7912 GATCTGGGTGATAGTAGTAGAGAAC-----GATATGTTAGCAATTTGGTGGTAC 7962
Db 1446 GluIleGluAsnIleLysThrAsnThrLysGluValThrMetLeuValSerAspSer 1465
QY 7963 TTACCTCCCGTAGAAATTAATATATATAGTTTTGAATACATATGAAATGAGAGAG 8022
Db 1466 LysSerAlaThrGluPheLys----- 1472
QY 8023 AATTAATATAGGTATACAAATTAATACAGATGTACAAACGTTTCTGCTGCTGG 8082
Db 1473 -----AspLysLeuSerGluCysLysAspLysGlyIlePheLysGlyIle----- 1487
QY 8083 GATGCTAATAGAAAGATATTTGAAAGCAATGACGTGCAGAACCCAGAGATGCAAAA 8142
Db 1488 -----ArgLysAspGluTrpGluCys----- 1494
8143 CTTTTAGAAAAGAGAAATGATGATTTGAACGCAATACATTA-----ATACAGATTAAG 8199
Db 1495 -----GlyLysValCysGlyValAspIleCysAsnLeuLysLysAspAsn 1510
QY 8200 TGTGACATTAAGAGCAGTACCTGTTGATGATTTATATATATCCCAACCGTTTGCATGATG 8259
Db 1511 IleGlyLysGluSerAsp-----LysLysTyrIleIleMetLys-----GluLeuLeu 1526
QY 8260 ACTGAATGCTCGATATATATTTGTAAGCACTGATGAGAGAAATTTGAAAAATTTAAAAA 8319
Db 1527 LysArgGlyTrpLeuGluLysPhe-----LeuGluAspTyrAsnLysIleLysHis 1542
QY 8320 TCATGTCAGTCGTAAAAACATCTGACAGATGCCAAGATGATTAAGTAAGTAAGTGT 8379
Db 1543 LysIleSerHisCysThrLysAsnGlyLys-----GlySerLysCys 1556
QY 8380 GAACAGTGTAAAAAGAGATGTCAGAAATATATAAATTTTCTTAAATGAAATCTCTA 8439
Db 1557 -----IleLysGlyCysValAspLysTrp----- 1564
QY 8440 TTGATATATGATCAATAATATACAAAGATTTGATGAACCAACAAATATATACAAAAATC 8499
Db 1564 ----- 1564

QY 8500 TCTACTATATGATCATGTTGCAAAATTTTGTACAAAAGTTGAAAACCTTTTAAGTGAATGT 8559
Db 1565 -----ValGlnGluInlYsGluGluInlYrPysGlnIleLys----- 1576
QY 8560 TCTGTGAGAGCTTTCTTCTGAAATATCTTCATGAAACAAAGTAAGTGTGAATTAATTT 8619
Db 1577 -----GluArgPheAsnGlu-----GlnTyrLysSer 1585
QY 8620 AATGAAATATGATGTTCTTCCATATATACAAATATGCAATATGCTTTCGAGAAACCAAAAAT 8679
Db 1586 LysThrSerAspGluLysPheAsnValLysSerPheLeuGluThrTrpIleProLys----- 1604
QY 8680 TATTAAGAACCTTGACAGTGTACACTACCTTTAAGAAATCCATGTAATTTGCTTACC 8739
Db 1604 ----- 1604
QY 8740 GATCAAAACAAAGATGATGATTAAGAAATTAACAACCTTTTACCTTCTGCGCAAGATGAT 8799
Db 1605 -----IleAlaValAlaAsnAsp 1610
QY 8800 TATGATTAATATCTTATATATTTGAAAGCATACCTGTTCTTAATATAGTTACAGATGATAC 8859
Db 1611 GluAspAsn-----ValIleLysLeuSerLysPheGlyAsnSer 1623
QY 8860 AAAGGTGATGATGTTCTCCCAAGAGAACATTTATGTACAGAACCTATACATGATAT 8919
Db 1624 CysGly----- 8919
QY 8920 AATTTATAGAAAAGGTGATTAAGAAATTTTAAAAAAAACCTTCTTATGCTTGTGCT 8979
Db 1634 Asn-----GlyAsnGluGlu----- 1638
QY 8980 CAAGGCAATTTGTATGCTCAAAAATTAATTAATCGAAGAGAGTGTGCTTTGAGCAATG 9039
Db 1638 ----- 1638
QY 9040 AATATATAGTTAGCAGATTAATTCGATATATTAAGAAAGTGAATGATGACACTTCA 9099
Db 1639 -----AspAlaIleAspCysMet 1644
QY 9100 TTATCTGAAAAATTAATTAATTAATTTGAAACATCAATTAAGAACCGAAATCGTAAA 9159
Db 1645 IleLysLysLeuGluLysLysIleAspLysLysArgLysProGlyGluAsn----- 1662
QY 9160 ACATGTTGGGAAAAATTAATAGACGTACAGTATGCAACGCTATGTATGTATTAATTAAT 9219
Db 1662 ----- 1662
QY 9220 GCTACTTCAAAAAGTAACATTAATGATGAAGATGTGTCATTAACCAAGATGAAGAACT 9279
Db 1662 ----- 1662
QY 9280 AATCAGTTTCTTCTGTTGTTAATGATGAGCAAGCAACATGTAAGAAAAACAAT 9339
Db 1663 -----SerGlyGlnThrLys----- 1667
QY 9340 GTAAGTATTCATTAATAACAAAATGTCTCTGTTCAACGAGATATTTTGAAGGTCA 9399
Db 1668 -----Asn 1668
QY 9400 GAATTTATTAAGACAACTGATGTCAGAAATGATTAAGAAAAATATATTAAGCTGAATATA 9459
Db 1669 GluThrLeuThrHisPro----- 1674
QY 9460 TTGATTAATAATCAATGGAATAATCTAAATATTAATTAAGCAATTAAGATCAATCT 9519
Db 1674 ----- 1674
QY 9520 TCAGATTAATTAAGCAATAAACCATCTGAAGAAATGTTCAAGTCAATATTAATCAAAA 9579
Db 1674 ----- 1674
QY 9580 GATTCATATGCGCTTTGAGTTAATGATTAATTAATGAATAGTTACAGAGAACAAAAAT 9639

```

Db 1675 -----LeuaspValGlnaspGluaspGluProleugluGluThrGluGlu 1689
QY 9640 AATGAAAATGATGATCAAGAGTACTAAAAAATTATATCGGTATATAT---TTT 9696
Db 1690 AspPro-----ValGlyLysGlnHisProSerPheCysProPro 1702
QY 9697 GTTGAGATGAAACACACAAAATCATGTACTGATGGAATATAAAAGAAAGAG--- 9753
Db 1703 ValGlnaspLysLys-----LysGluGluGluGly 1712
QY 9754 CAACAGTTCGCTCAAGACACTATTTCTTACACCCCATGTAGTTCTTCTATCA 9813
Db 1713 GluThrCysThrProAlaSer-----ProAlaProAlaProAlaPro 1726
QY 9814 GCACCTTATTCATCAACACATCGATAGACAAATATGATCCTTAAATGAT-----ATA 9867
Db 1727 AlaProAlaSerProSerProThrProAlaProAlaaspLupProPheaspProThrIle 1746
QY 9868 TTGAAAAGTAGATCTCTGTTGTTATGTATCGCGCTTAGGTTGATAGCGCTTCATTC 9927
Db 1747 LeuGlnThrThrIleProleuGlyIleAlaLeuAlaLeuGlySerIleAlaPheLeuPhe 1766
QY 9928 ATGAG----- 9933
Db 1767 LeuLysValIleTyrIleCysValValTyrMetTyrIleTyrMetCysPheCysIleTyr 1786
QY 9934 -----AAAAATTCAAATGCTGTGAGACTGTGCGTATGATGATCCGCAAGA 9987
Db 1787 MetTyrLysLysThrLysHisProValAspLeuPheSerValIleasnIleProLysSer 1806
QY 9988 GAGTATGGAATGCCATCGTGTGGAATCCAAAATAGTATACCAATATAGAGTGTCCA 10047
Db 1807 AspTyrAspIleProThrLysLeuSerProAsnArgTyrIleProTyrThrSerGlyLys 1826
QY 10048 TATTAAGCGCAAAATATATATATATATGAGAGATCT-----ATGCGAGATGAA 10098
Db 1827 TTTATGGLysArgTyrIleTyrLeuGluGlyAspSerGlyThrAspSerGlyTyrThr 1846
QY 10099 GATAAATATATGTGCGACTTATCTTCATATATCT---TCATCCGAAAGAGATAT 10155
Db 1847 AspHisTyr-----SerAspIleThrSerSerSerGluSerGluTyr 1860
QY 10156 GAAGAATGATATTAATGATATATATATGACAGAGTATGCTTAAATATAAACATTGATA 10215
Db 1861 GlnGlnMetaspLysasnAspIleTyrValProGlySerProLysTyrLysThrLeuIle 1880
QY 10216 GAAGTAGTACTAGACCATCAAAAAGGATATACCAAGTATGATACCAAGT----- 10269
Db 1881 GlnValValLeuGluProSerGlyAsnAsnThrThrAlaSerAspThrGlnAsnAspIle 1900
QY 10270 ---AATGATACACACGATGAGATAGATATTTATGATGATGATGATGATGAT 10326
Db 1901 GlnAsnaspLysIleProSerAsnLysPheSerAspAsnGluThrLeuLysAsp 1920
QY 10327 GATTGTGATCTCAATTTTACCAATACGAACCAAT-----AATATTTACAAAGT 10380
Db 1921 AspPheIleSerAsnMetLeuGlnAsnGlnProLysaspValProAsnAspTyrLysSer 1940
QY 10381 GCAGATTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10440
Db 1941 GlnAspIleProPheAsnThrGlnProAsnThrLeuThrPheaspLysProGluGluLys 1960
QY 10441 CATTATATATATATATATATATATATATATATATATATATATATATATATATAT 10500
Db 1961 ProPheIleThrSerIleHisAspArgAsnLeuLeuasnGlyGluGluTyrSerTyrAsn 1980
QY 10501 ATTAATATGATCTATATATATATATATATATATATATATATATATATATATATAT 10560
Db 1981 ValAsnMetSerThrAsnSerMetAspAsp---ProLysTyrValSerAsnValValTyr 1999
QY 10561 AGAGGTATGATTTATATATATATATATATATATATATATATATATATATATATAT 10619

```

```

Db 2000 SerGlyIleAspLeuIleAsnAspSerLeuSerGlyAsnLysHisIleAspIleTyrAsp 2019
QY 10620 GAAGTATTTG 10628
Db 2020 GlnValLeu 2022

RESULT 11
T28634
Variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28634
R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfelde, J.A.; Peterson,
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: 220487, PMID:95330813, PMID:7606788
A:Accession: T28634
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2182 <SD>
A:Cross-references: EMBL:LA2636; NID:g886379; PID:g886380; PIDN:AA75399.1
C:Genetics:
A>Note: var-7

Alignment Scores:
Pred. No.: 5,33e-79 Length: 2182
Score: 1755.00 Matches: 760
Percent Similarity: 29.37% Conservative: 337
Best Local Similarity: 20.35% Mismatches: 678
Query Match: 9.02% Indels: 1960
DB: 2 Gaps: 125

US-10-087-013-1 (1-10628) x T28634 (1-2182)
QY 37 GGGAAATCGACATCATATAGAGGAGATGCTTAAAGCCCTATATTAAGAAAGATCAC 96
Db 10 GlySerIleGlySerSer---SerGlyLysGlyLysLysAspThrSerGluTyrIleTyr 28
QY 97 AAAAGT---GCAAGAAATGTTTGAACGT-----TATGCCAAAATATA 138
Db 29 ValSeraspLysLysAspLeuLeuaspArgValGlyGluLysValTyrGluGluLysVal 48
QY 139 AGACATCCATCAAAATATGCAAAAAGAACATGTGATTCGTTGAAAGGATTTGACGAAA 198
Db 49 LysasnGlyasp-----AlaLysLysTyrIleGluAlaLeuLysGlyAsnLeuAsnThr 66
QY 199 GCAGAAATTTGCTGTGGTCCCTACGCCAGTAAATAGCATATATATATATATATATAT 258
Db 67 AlaasnGlyAspSerSerGluThrAlaSerSerIleGlu----- 79
QY 259 CCATGTAATTTAGATCATTAAGAACATACATTAATTTAGCGTATGATGATGAT----- 312
Db 80 ThrCysThrLeu---ValLysGluTyr-----TyrGluArgValAsnGlyasp 94
QY 313 ---TTGAGACATCTTGCCAT-----GTAGAGAACAAAACCATTTGATGAA 357
Db 95 GlyLysArgHisProCysArgLysAspAlaLysasnGluaspValAsnArgPheSerAsp 114
QY 358 GATGAGAAATCTGAATGATGCA---AATAAATATGATATATTAAGAAAAAATGATGCT 414
Db 115 ThrLeuGlyGlyGlnCysThrTyrAsnArgIleLysaspSerGlnGlnGlyAspAsnLys 134
QY 415 ATA---GCTGTGGCGCACCTAGAGAGACATATGCTGATTAATAAACTTGAAGCTCTA 471
Db 135 ValGlyLysAlaProTyrArgArgLeuHisLeuCysaspTyrAsnLeuGluSerIle 154
QY 472 AATGATATTAATACCAAAATATCATGATTTATGGGAAATGACTACTACGCAAAA 531
Db 155 Asp-----ThrThrSerThrThrHisLysLeuLeuGluValCysMetAlaLys 172
QY 532 TACGAGGTGATCAATGATTTGTAATATATATATATATATATATATATATATATATATAT 576
Db 173 TyrGluGlyAsnSerIle---AsnThrHisTyrThrGlnHisGlnArgThrAsnGluasp 191

```

QY 577 -----TCAGACGCTGCTCTCTGACGAGAGTTTGCAGATATAGTGATATGTA 630
 Db 192 SerAlaSerGlnLeuGlyThrValLeuAlaArgSerPheAlaAspIleGlyAspIleVal 211
 QY 631 AGAGAAATGAGAAATGTT-----AAACCAAAATGTCATGACAAAGTA 672
 Db 212 ArgGlyLysAspLeuGlyLeuGlyTyrAspAsnLysGlnLysGlnArgLysLeu 231
 QY 673 GAACGGGCTCCGAGAGTTTTCAGAAATACATGATGAAAGAGAGTGAAGTAA 732
 Db 232 GlnGlnLysLeuLysAspIlePheLysLysIleHis-----LysAspValMetLys 248
 QY 733 AATGAT-----TACAAATCCTATGGA---TCTGAAATATATATATATTA 774
 Db 249 ThrAsnGlyAlaGlnGlnLysArgTyrIleAspAspAlaLysGlnLysAspPheGlnLeu 268
 QY 775 AGAGAAAGATGAGTGAATGATGAGAAATTAAGTATGAGAGTATTAACATGATGCA 834
 Db 269 ArgGlnAspTrpTrpThrSerAsnArgGlnThrValTrpLysAlaLeuIleCysHisAla 288
 QY 835 TCATATATATCTGATATTTTATGCAATCAGAAAGTATACACCATTTATTTCAATCT 894
 Db 289 ProLysGlnAlaAsnTyrPheIleLysThrAlaCysAsnValGlyLysGlyThrAsnGly 308
 QY 895 AATGCGGCCATTAACAAGAAAGAGTCCATCCAAATTAATATGTCCTCAATATTA 954
 Db 309 GlnCysHisCysIleGlyGlyAspValProThrLysPheAspTyrValProGlnTyrLeu 328
 QY 955 CGTTGGTTCAGAGATGGGAGAGAGTTTGCAGAAAGAAATTAATTAATGAAAG 1014
 Db 329 ArgTrpPheGlnGlnLysArgPheCysArgLysLysLysLysLysLysLys 348
 QY 1015 GTCAAGACATCTGCTGATGACAAAGACCGTATATTTGATGCAATTAATGACATGAT 1074
 Db 349 LeuGlnLysGlnCysAsnArgAspTyrGlnGlnAsnLeuLysCysSerGlnLysTyrAsp 368
 QY 1075 TGATGCAACATATTTGAGAAAGAGTATTTGATTTGATTAATAGTACTGACTGT 1134
 Db 369 CysThrLysThrIleLysLysGlyLysLeuValIleGlyLysHisCysThrAsnCys 388
 QY 1135 TGCATTAATGCAAGTTTGTGAGTTGGTATGAGATCAACAGACATTTAAATA 1194
 Db 389 SerValTrpCysArgMetLysGlnTrpIleAspAsnGlnLysGlnPheLeuLys 408
 QY 1195 CAAAGAAATATATGAAAGAAATA----- 1221
 Db 409 GlnLysArgLysTyrGlnTrpIleSerGlyGlyGlySerGlyLysSerProLysArg 428
 1222 -----CAATCAATTTATGACAGATACAAATTTGTCATATAT 1263
 Db 429 ThrLysArgAlaAlaArgSerSerSerSerSerSerSerSerSerSerSer 442
 QY 1264 AATAGAT 1323
 Db 443 -----GlyTyrGlnSerLysPheTyrLysLysLysLysLysLysLysLysLys 460
 QY 1324 GACACTTTTAAATTTACTAAATGAGAAAGTATTTGTAAGA-----GGA 1371
 Db 461 AspLysPheLeuLysIleLeuAsnLysGlnGlyLysGlnLysGlnProGlnValCys 480
 QY 1372 TTACAGAGGAAAGAGATTTACTTTACTTAACAGTGTGATGACAAAGAGATTTTAT 1431
 Db 481 AsnGlnLysAlaAspAsnValAspPheThrAsnGlnLysTyrValLys---ThrPheSer 499
 QY 1432 CGTTGAGATATTTGCAAGTGTGCTCCGACAGCGGGGCAATGATGATGATTAATAC 1491
 Db 500 ArgThrGlnIleCysGlnProCysProTrpCysGlnGlnLysGlnValGlyProTrp 519
 QY 1492 ACACCAATATGATATGATGATGATGATGATGATGATGATGATGATGATGATG 1551
 Db 520 LysValLysGlnAspLysThrCysGlySerAlaLysThrLysThrLysAspProLys--- 538

QY 1552 GGTGTGAACCTACTAATATCACTGTCTTATAGTGTATGACAAAGGTGATTTACA 1611
 Db 539 -----AsnIleThrAspIleProValLeuTyrProAspLysSerGlnGlnAsnIleLeu 556
 QY 1612 CAAAATTTAGAAATTTTGTACACAGCTCACTCAATATTAACAAGATTAATAATCAAAA 1671
 Db 557 LysAlaTyrLysAsnPheCysGlnLysGlyAlaPro---GlyGlyGlnIleLysLys 575
 QY 1672 TGGGAATGCTATATATAGAT-----GAAATATATATATGATGATTAACAG 1716
 Db 576 TrpGlnCysTyrTrpArgPheGlnHisArgProSerSerLysAsnAsnAsnLysValGln 595
 QY 1717 GAACAAATATCTAATCAATATATATATATATATATATATATATATATATATATAT 1776
 Db 596 GlyThrTrpAspLysPheThrGlnGlyLysGlnThrValLysSerTyrAsnValPhePhe 615
 QY 1777 GAATATGGGTATACATATATATATATATATATATATATATATATATATATAT 1836
 Db 616 TrpAspTrpValHisAspMetLeuHisAspSerValGlnTrpLysThrGlnLeuSerLys 635
 QY 1837 TGTATTAATATATACAAACACG-----CATGTATGATGAATGTAACAGAAAT 1884
 Db 636 CysIleAsnAsnAsnThrAsnGlnLysAsnThrCysArgAsnAsnAsnLysCysLysThrAsp 655
 QY 1885 TCGTTATGTTTTCAGAGATGGGTAAACAAAGAAAGAAAGATGATATATTAAGAA 1944
 Db 656 CysGlyCysPheGlnLysTrpValGlnLysLysGlnGlnLysTrpMetAlaIleLysAsp 675
 QY 1945 CTGTTCACAAAAGAAAGAAATATACAGCAATCGTATATATATATATATATATAT 2004
 Db 676 HisPheGlyLysGlnThrAspIleValGlnGln-----LysGlyLeuIle 690
 QY 2005 GAAGTATATTTTAAATGATATATATATATATATATATATATATATATATATAT 2064
 Db 691 ValPheSerProTyrGlyValLeuAspLeuVal-----LeuLysGlyLys 706
 QY 2065 CTTATGAAAT 2124
 Db 707 LeuLeuGlnAsnIleLysAspValHisGlyAspHisAspIleLysHisIleLysLys 726
 QY 2125 TATTTAGAG-----AATGCAATA 2142
 Db 727 LeuLeuAspGlnGlnAspAlaValAlaValLeuGlyLysAspAsnThrThrIle 746
 QY 2143 GAACCTGTTAGATCACTTAAAGAAAGCCGATGATGATGATGATGATGATGATGAT 2202
 Db 747 AspLysLeuLeuGlnHisGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 764
 QY 2203 GAACATGTTGAACATCCCATATGCAACAAACCCGTTGTTAAACCTCGTGAGGC 2262
 Db 765 GlnLysCysGln----- 768
 QY 2263 ACGCAACCACTAATAATATATATATATATATATATATATATATATATATATAT 2322
 Db 769 -----LysLysAlaGlnGln 773
 QY 2323 GAACACGAAATGCTGCTTCAATATTAAGAAAGACACAGAGTATATATAA 2382
 Db 774 GluSerArgLysArgSerAlaGln----- 781
 QY 2383 CGTGGGGGTGAGAAAGCACTCAAGGACATTTATGTAATATATATATATATATAT 2442
 Db 782 -----ThrArgGlnLysGlnArg 787
 QY 2443 AATGTAATCTTGTTTTCATATGACCATGATGATGATGATGATGATGATGATGAT 2502
 Db 788 -----ThrGlnGlnProAlaAspSerAlaGlyGln----- 797
 QY 2503 CAAACAAGATTTGCTAGAACCTGAATGGAATGATCCGGAACACATGCTGAAGAT 2562
 Db 798 -----ValGlnGlnGlnLysAspAspAspAspAspAspAspAspAspAsp 810
 QY 2563 CACGAGATGTTATATATGCTCTGAGAAAGACGACATATATATATATATATATAT 2622

Db 811 ASPGLUASP----- 813
 QY 2623 TTACAAACGGATGATCACCACCTTAATGTAAATTGTTGATGATTAGTAAATATTC 2682
 Db 814 -----
 QY 2683 TTTTGGGGGATGTTCTTCTTATCAGCAAAATVANGAACCAAGATTAATACGAATGTAT 2742
 Db 821 GLUGLUGLY----- 823
 QY 2743 AAGAAAGAAATAACCTTAAGGGCCCCAAGAGTAACGCCAACACAGACACT 2802
 Db 824 LysGLUGLUGLYThrValIhrGluValIhrGluValIhrGluValIhrGluIuThr--- 842
 QY 2803 ATCTGCGACCTATAGCTTACAGTTTGCAGATATAGTATTAATTCGAGNAGAGAT 2862
 Db 842 ----- 842
 QY 2863 CTCTGGGAAAGAAACGGTCATGTAAAGCTGCAGACACTTGGAACTGTTTTGGT 2922
 Db 842 ----- 842
 QY 2923 AATATACATAGTCACTCAAGGCAAGCAAAATGATTAATATATGATGCCCCCAA 2982
 Db 842 ----- 842
 QY 2983 TATTTAAATTGAGGAAATTTGGTGGAACTAATAGACCAAAAGTATGGAAAGCATG 3042
 Db 842 ----- 842
 QY 3043 AAATGATATTAATAATTTTGAAGATAAATCGGAGCACCAATCAACAAAGTAGTTAT 3102
 Db 842 ----- 842
 QY 3103 TGGGATATAGTATCATCATACACATTCGATGATTAATATCCACAAAATTAAGATGATG 3162
 Db 843 -----Val 843
 QY 3163 ACCGAATGGCGAGATGTTACTGTCAGAGTGCAGAAAAAGAGTATGTAAGTTGAGAGAG 3222
 Db 844 ThrGlu-----GluGlu 847
 QY 3223 AATGTAAAGAGTAAAGATTAAGATTAATGTCAGAGCTGTACGAAAGAGAGTGTACA 3282
 Db 848 GlyValLysProCys----- 852
 QY 3283 GGTGTACGAAGTGCACAGAGCTTGTAAATGATTAATATAGATTAATAGATTAAGGAAA 3342
 Db 853 -----AspIleValIglyLysLeu--- 858
 QY 3343 GAACAATGAATATATATATCAGATTAATACAAAGATTAATCATGACAAAGCAAAATGTCT 3402
 Db 858 ----- 858
 QY 3403 GTTAGTAATAGTGTATGTAGAGCTCCAGTACTGCCAAAATCATATAGACAGAAATGTT 3462
 Db 858 ----- 858
 QY 3463 ATTGAATTTTTGTGCAATTAATACAACAAAATGTGTGCAAAAGTAAATAAGTGTACT 3522
 Db 858 ----- 858
 QY 3523 AGTGAGAAAGTGTCTCATTTGTTACTACACACGATGATAAATGTTGAGCATATATCTC 3582
 Db 858 ----- 858
 QY 3583 CATGATACAGAAATTTTGTATGATGTGCATCACAATAATGAGTTTGTGATGAAAAAGT 3642
 Db 859 -----PheGluAspAspLysSerLeuLysGlu 867
 QY 3643 GATGTAAAGATTAACGAAAAATATGCTTATAGAGATTAACACAGACCATGATGTGCG 3702
 Db 868 -----Ala 868
 QY 3703 TGTGTTGTAATAAGTGCATGCAAAACGCAAGGTTACAGATTAATAACGAAAAAAGCG 3762
 Db 869 CysGlyLeuLysTyrGlyPro----- 875
 QY 3763 GAAGAAAAAGATACGGAATGTAAAAACAGTAAATATTAATAAGAAAAAGATGAAG 3822
 Db 876 -----GlyGlyLys 878
 QY 3823 AAACAAGTAAAGATTTGTATCCAAAAAAGATTAATGATATCCGATTTGGCATATGC 3882
 Db 879 GluLys-----PheProAsnTyrLysCys 886
 QY 3883 GGAATATTAATTAAGT-----GAGAACCTCTGTGTATGCC 3924
 Db 887 ValIhrProSerGlyValSerThrAlaThrSerGlyAspGlyAlaIleCysValPro 906
 QY 3925 CTRAGAACCAAAAGTTATGCGTACATTTCTTGCA-----AATGATAAAT 3969
 Db 907 ProIhrGhrIhrGhrIhrGlyValGlyLysLeuSerIhrProLysSerArgGlyLysP 926
 QY 3970 GAATATAAAAAATTACATCAACAAAGT-----AAT 3999
 Db 927 GluIhrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGluLys 946
 QY 4000 TTAAGAGACCTTTCATCAATCTGCAGACAGCAAAACATTTCTCATGTATATATAT 4059
 Db 947 LeuIhrGhrIhrAlaPheIleGlnSerAlaIleIleIhrPheLeuIhrPheLysTyr 966
 QY 4060 AAA-----AGTAAGAGTCT----- 4074
 Db 967 LysGluGluLysLysProProAlaThrGlnAspLysIleGlyLeuGlyValSerLeuPro 986
 QY 4075 -----GAGGAAATGTAACCTGATTAAGAAATTAAGAA-----GGCAAAATTCCT 4119
 Db 987 GluProSerProProGlyGluAspProGlnThrGlnLeuGlnIhrGlyValIlePro 1006
 QY 4120 CCGCATTTTGTGATCATCATGTTCTTACACATTTGGAGATTAATAGATTTTATTTGGA 4179
 Db 1007 ProAspPheLeuIhrGlnMetPheTyrThrLeuIhrAspLysLysAspIleLeu----- 1024
 QY 4180 ACAGATATATCAAAAGTCAATGTGTAGAGGAATTAACCTAAAGCAAAATATGATTCCTT 4239
 Db 1024 ----- 1024
 QY 4240 TTCAAAAATGTGACCAAAAATCTCTAATGAAAAACGCCAAAGATGTGACAGAA 4299
 Db 1024 ----- 1024
 QY 4300 CATATGATGATGATGGAAGCTATGTGCTATGTGCTAATAAATTTGGGGCAAAAAA 4359
 Db 1024 ----- 1024
 QY 4360 GATGATTTTACCGAAACTACGGTTACACACAGCTCAAAATTTAGTGACAAAAAGCACACT 4419
 Db 1024 ----- 1024
 QY 4420 TTGAGGAATTTGCCAAAGACCCAGTTTTCAGATGCTAACCGAATGTGACAGAC 4479
 Db 1024 ----- 1024
 QY 4480 TATGCTATACACGCAAAAAATATTTGAAGATGTGACGAAAAATATGATCAATGAC 4539
 Db 1024 ----- 1024
 QY 4540 CAATGAAGTGTGTACAGATGTAAATAGAAATGCCAGACTACGTTAAATATATGAAA 4599
 Db 1024 ----- 1024
 QY 4600 AAAAAAAGAGTGCATTCACAAGATTAATTAACAAGATGAACGCCGACAAAAAAGA 4659
 Db 1024 ----- 1024

Db 868 -----Ala 868
 QY 3703 TGTGTTGTAATAAGTGCATGCAAAACGCAAGGTTACAGATTAATAACGAAAAAAGCG 3762
 Db 869 CysGlyLeuLysTyrGlyPro----- 875
 QY 3763 GAAGAAAAAGATACGGAATGTAAAAACAGTAAATATTAATAAGAAAAAGATGAAG 3822
 Db 876 -----GlyGlyLys 878
 QY 3823 AAACAAGTAAAGATTTGTATCCAAAAAAGATTAATGATATCCGATTTGGCATATGC 3882
 Db 879 GluLys-----PheProAsnTyrLysCys 886
 QY 3883 GGAATATTAATTAAGT-----GAGAACCTCTGTGTATGCC 3924
 Db 887 ValIhrProSerGlyValSerThrAlaThrSerGlyAspGlyAlaIleCysValPro 906
 QY 3925 CTRAGAACCAAAAGTTATGCGTACATTTCTTGCA-----AATGATAAAT 3969
 Db 907 ProIhrGhrIhrGhrIhrGlyValGlyLysLeuSerIhrProLysSerArgGlyLysP 926
 QY 3970 GAATATAAAAAATTACATCAACAAAGT-----AAT 3999
 Db 927 GluIhrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGluLys 946
 QY 4000 TTAAGAGACCTTTCATCAATCTGCAGACAGCAAAACATTTCTCATGTATATATAT 4059
 Db 947 LeuIhrGhrIhrAlaPheIleGlnSerAlaIleIleIhrPheLeuIhrPheLysTyr 966
 QY 4060 AAA-----AGTAAGAGTCT----- 4074
 Db 967 LysGluGluLysLysProProAlaThrGlnAspLysIleGlyLeuGlyValSerLeuPro 986
 QY 4075 -----GAGGAAATGTAACCTGATTAAGAAATTAAGAA-----GGCAAAATTCCT 4119
 Db 987 GluProSerProProGlyGluAspProGlnThrGlnLeuGlnIhrGlyValIlePro 1006
 QY 4120 CCGCATTTTGTGATCATCATGTTCTTACACATTTGGAGATTAATAGATTTTATTTGGA 4179
 Db 1007 ProAspPheLeuIhrGlnMetPheTyrThrLeuIhrAspLysLysAspIleLeu----- 1024
 QY 4180 ACAGATATATCAAAAGTCAATGTGTAGAGGAATTAACCTAAAGCAAAATATGATTCCTT 4239
 Db 1024 ----- 1024
 QY 4240 TTCAAAAATGTGACCAAAAATCTCTAATGAAAAACGCCAAAGATGTGACAGAA 4299
 Db 1024 ----- 1024
 QY 4300 CATATGATGATGATGGAAGCTATGTGCTATGTGCTAATAAATTTGGGGCAAAAAA 4359
 Db 1024 ----- 1024
 QY 4360 GATGATTTTACCGAAACTACGGTTACACACAGCTCAAAATTTAGTGACAAAAAGCACACT 4419
 Db 1024 ----- 1024
 QY 4420 TTGAGGAATTTGCCAAAGACCCAGTTTTCAGATGCTAACCGAATGTGACAGAC 4479
 Db 1024 ----- 1024
 QY 4480 TATGCTATACACGCAAAAAATATTTGAAGATGTGACGAAAAATATGATCAATGAC 4539
 Db 1024 ----- 1024
 QY 4540 CAATGAAGTGTGTACAGATGTAAATAGAAATGCCAGACTACGTTAAATATATGAAA 4599
 Db 1024 ----- 1024
 QY 4600 AAAAAAAGAGTGCATTCACAAGATTAATTAACAAGATGAACGCCGACAAAAAAGA 4659
 Db 1024 ----- 1024

```
OY 4660 TTGATAGACAACACATGTTGTTAAGTTAGACATCTACTGGAAGCATGCAACAGAT 4719
Db 1025 -----
OY 4720 TACTTGAACAGGAATTTACTGCTTACTGTTGTTGATTAAGCCTGGAAGTGCCTGTGTGTA 4779
Db 1032 -----
OY 4780 CAAAGAAATATACATTTGTTAGAAAACAGGCTTACTATGATGCCGACAAACATTTGTGG 4839
Db 1032 -----
OY 4840 TGCACAAAATTTATTTGAAAATGACGACAAATATCTAACAATTCGAGTAAAGATAGTGC 4899
Db 1033 -----
OY 4900 AAAGATTAGTAAAGGAGGCAACACAGGTGCTATTAGTGGCAAAAACAAAGTCTTAAT 4959
Db 1048 -----
OY 4960 AACTACAACTTGAAGAATTGACTGAAGATGCTTTTCTCTGTCGACTACGT 5019
Db 1055 -----
OY 5020 ATATGTTTTCATGATTCGATTCGCAATTATACAGATCCAGAGTTAAAGATGAATAATGGG 5079
Db 1055 -----
OY 5080 TTGCGAAAAGATTTGATGGAAGTGGGGCAAGGAGGTACAAATTTGGTCAATCTAC 5139
Db 1056 -----
OY 5140 AAAAGAAAAAAGAAAAAGAAAAATTAACATCGATGCGACCAAAATTTCTTATGAG 5199
Db 1060 -----
OY 5200 GTCCCGCTTGTAGTGTATGAATAATAGTTTATGATTTAAGACATATATTTAGGT 5259
Db 1066 -----
OY 5260 ATTGATTAATTTGAGATGAGAAAAACAAAGACGAGAAAAATTTGAGAAAAATTTTAC 5319
Db 1067 -----
OY 5320 AAAAATGAAACATCAGTTGGCAAGAGATGATAGTACAGGAATCCCGTACTACT 5379
Db 1078 -----
OY 5380 GCGCGAAAAATTTTCTGGAACGAAAAATTAAGATGTTGTGGAAGCATGATGCGGG 5439
Db 1094 -----
OY 5440 -----TACAAACGTGTGAGGATGAGTAAGTAAGTGAAGTGAAGATGATGCA 5493
Db 1114 -----
OY 5494 GATCTAAAAAAATGTGGTCTCTACTTCAATGATGATTAATCTATGCGGCAAAAATGCG 5553
Db 1122 -----
OY 5554 GATGAGTACTGCGTACATCTTCTGAGTGTTCGCGATGGGGAAGATTTTTCG 5613
Db 1123 -----
OY 5614 AAACATTAAGAAAGAAATTTGCAAAATTTGAGGCGCTGAATGATTAATCTGTGT 5673
Db 1125 -----
OY 5674 GATTAATGAATTAAGAAAGAAATGTACAGATGCGTGAACAATATTAATAATTTAT 5733
Db 1125 -----

OY 5734 AGTAGTGAAGAACACAGTATGAAAAACAAATCAAAAAATATGCTGAGATTAAGACAAA 5793
Db 1126 -----
OY 5794 ATATATTCGAGCATCCTGTGGCAAAAAGATGACAGAGAGCTGCGCAATATTTAGCAAAA 5853
Db 1133 -----
OY 5854 CAATTAATAAAATTTGTGAAATTAAGTGAATTTGTAATTAATGATGAAGAT 5913
Db 1134 -----
OY 5914 GTGTCAACAGCATTAATGATGATTAATGATCAAAATATCCGCGATCATTAAGCAT 5973
Db 1137 -----
OY 5974 GAACCAAAAAGTTGGAAGAAAGTAAATTTGTAAGTCCACGAGTCCACCGAGTGA 6033
Db 1147 -----
OY 6034 CGAAGGAAACACCGTCCACGAGGATATCATGATATCAAAAACGAGCATGAAAAA 6093
Db 1147 -----
OY 6094 GAAGCAAAAACGCGCGCTTACAAAACGCGGAAAAAGTGAATTAACAACAGAA 6153
Db 1147 -----
OY 6154 ATGCGAGCAAAACAGAACCGGACGAGCAGCAACAAACGAAACGAAATCAACA 6213
Db 1147 -----
OY 6214 GCAACACACAGATCTGACGTGGGCAATGTTAAAGCCATTTCTGCAATTAACA 6273
Db 1147 -----
OY 6274 GATACAGGGGTGAATAGAGGTTGTAATCCAAAACGTATGACATATCTAAATG 6333
Db 1147 -----
OY 6334 GGTGTATTTAGTATGATCTTAAGAAATTAAGAAATGAGCATATGATGCTCTAGAGA 6393
Db 1147 -----
OY 6394 AAAAATTTGTAATTAATATATATCAATATTTAATTAATGAACTGAAATTAACGTGAC 6453
Db 1148 -----
OY 6454 AATGATTAATAAGAGGCTTTTATTAATGTCACCAATAGAACTCAATTTTGTGTTA 6513
Db 1159 -----
OY 6514 AAATATTAATTTGAAATCTCGACGAGAAATTAATGCAAAAATGAGCAAAATTCAGAT 6573
Db 1159 -----
OY 6574 GAATTAATAAGAAATGATTAATATACATATGTCATTAATAAGATATTTTGGACT 6633
Db 1159 -----
OY 6634 GATATTTCTAATGATTAATAAAATTAATAGCTGAACAATAGTGAACACCATTCAT 6693
Db 1160 -----
OY 6694 GAAATATTAAGAAACAGATTAATAAAATTAAGTAAATTAAGTAAATTTTGG 6753
Db 1161 -----
OY 6754 GAGAAAAATTAATAATTTATTTGGAGAGATATATGATTAATTAATCAATCA 6813
Db 1168 -----
OY 6814 GACGAAAAAGAAAAAATTTAGAGATTAATTAACAGTATGACATGACCAAAATG 6873
```

[illegible]

D	b	1385	AsnGlyasnIleGluMetValIalSerAspSerThrAsn-----	1398
O	y	7882	ATATGTGATACACTATGAATAATAGTTTCCTGGTGCATCTGGTGACATAGTTAGAGAAACAGAT	7941
D	b	1399	-----ThrpheGlnHisLeuGlyAsp-----	1405
O	y	7942	ATGTTCGAATGTGGTGTACTTACTTACCCTCCCGTAGAATAATAATTATATAGCTTTTGCA	8001
D	b	1406	-----CysLysSerSerylIlePheLys	1413
O	y	8002	TACATATATGGAAAAATGGAGAATAAAAATTAAGTAGAAAAATTAATACACGATGTACAA	8061
D	b	1414	GlyIle-----	1415
O	y	8062	ACGTTTCGTTCTGCTGGTGGATGCTTAATAGAAAAGATATTGGAAA--GCATAGCG	8118
D	b	1416	-----ArgLysAspGluTrpLysCysAlaAsnVal	1425
O	y	8119	TGCAAAGCACCAAGATGCAAACTTTTACAAAAAGAAAGATGGATGGATTGACCC	8178
D	b	1426	Cys-----GlyValAspIle	1430
O	y	8179	ATAACATTAAATACAAGAT--AAGTGGACATAGAGCATCCACCTGTGATGATAT	8235
D	b	1431	CysThrLeuGlnLysLysIleLysAsnGlnGlnGlyAsp-----LysLysTy	1447
O	y	8236	ATACCTCAACGGTTTCGATGGATGACTGAATGGCTCGAATATTATTTAGAACCTGATG	8295
D	b	1448	IleThrMetLys---GluLeuLeuLysArgTrpLeuGlnTyPhe-----Leu	1462
O	y	8296	CAACAATTTGCAAAAATTTAAAAAAATCATGTGATCACTG---AAACATCTGCACGATGC	8352
D	b	1463	GlnAspTryAsnArgIleArgLysLysIleLysLeuCysThrLysLysGlnLysPelLys	1482
O	y	8353	AAGATGATTATGATGA AAAATTAAGTGTGAACAGCTGAAAACAGATGCAAGCAAAATATAA	8412
D	b	1483	Lys-----CysIleLysGlyCysIleGlu-----	1490
O	y	8413	AATTTTGTCTTAANTGAAATCTCTATTCGATATACATCAAAAATAATACAAAGATTG	8472
D	b	1491	-----LysTrp-----	1492
O	y	8473	TATGACACCAACATATATACAAAAATCTCTACTTATGATCATGTTCAAATTTGTACAA	8532
D	b	1493	-----ValGln-----Glu	1495
O	y	8533	AAGTTGAAACTTTTAAAGTGAATGTCTGTGTGAAGCTTTTCGAATATCTTCATGAA	8592
D	b	1496	LysThrLysGluTrpGln-----	1501
O	y	8593	ACAAGTAAGTGTGATTAATTAATTAATGAANAATGATGTTCTCCAAATATAGACAA	8652
D	b	1502	-----LysIleAsnAsp-----Thr	1506
O	y	8653	TATGCTTTGCAAGAAACACCAAAAAAGTTATTAAGAAAGCTTGACGTTGACACTACCTTC	8712
D	b	1507	Tyr-----	1507
O	y	8713	AAGATCATGTGATTAATGTGCTCACGATCAAAAACAAAAGATGATGTAAGAAATTACAA	8772
D	b	1508	-----LeuGlu-----	1509
O	y	8773	ACTTTTACCTCTGCTCGAAGATGATTAATATTAATCTTGATTAATGGAAGCATAC	8832
D	b	1510	GlnTy-----LysAsnAspAspGlyAsnThrLeuThrAsnPhe-----	1522
O	y	8833	CTTGTTTATATAGTTCAGATGATTAACAAAGGTGATTAATTCCTCCACAGAAGACAT	8892
D	b	1522	-----	1522
O	y	8893	TTATGTACAGACTATACACTGCATATTAATTAAGAAAAAGTGTATRAAGAAATTTTAAA	8952
D	b	1533	-----LeuGlnGlnPheGlnTyArg-----	1539

QY 8953 AAAAAGCTTCTTCTGCTTTCAGTCACAGCAATGTAGGTCAAAAATATAATATCG 9012
 Db 1529 -----
 QY 9013 GAAGAAGAGTTGCTTTGAGGCAATGAATATAGTATGACAGTTATTCGATATTAAT 9072
 Db 1530 -----
 QY 9073 AAAGCACTGATATGATGACACCTTCATATTCGAAAAAATTAAAAAATATTGAACA 9132
 Db 1537 LysProCysaspGlyLeuasp-----ThrgluPheLysAsnAlaIle 1536
 QY 9133 TCAAAATGACACCAAGCAAAATCGTAAACATGCTGGAAAAATATAGACGTAGATATG 9192
 Db 1543 -----
 QY 9193 CAGCATATGATATGATATTAATATGCTACTCAAAAGTAACTATGATGAAGATGG 9252
 Db 1543 -----
 QY 9253 TGTCAATTCACCAAGATGAGAAACATCACTGTTCTGCTGTTAATTGAATGGCA 9312
 Db 1544 -----
 QY 9313 AAGCAAGCATGTAGAGAAAAAGAAACATGTAGATGATTCATTAACCAAAATGCTCGT 9372
 Db 1546 -----
 QY 9373 TCAACGAGATATATTTTGAAGCGTCAGATATATTAAGACACCTGATGATGATGAT 9432
 Db 1552 AsnSerThrAspAsn-----SerGlnAsnLysAsnAsnAsp 1564
 QY 9433 ATTAGAAAATATATTTAGCTTGAATATATGATTAATAAATACATGAAATCTAAATATA 9492
 Db 1565 -----
 QY 9493 AATATAGCAATTAAGATCAATCTCAGGTAATATAGACATATACCATCTGAAGA 9552
 Db 1576 LysIleSerGlnCysLysGlnGlnHisSerGlyGlnThr--ProCysasp--- 1593
 QY 9553 AATGTCTACATATATTAATCAAAAGATCTCAATGCGCTTGGATTAATGATATA 9612
 Db 1594 -----
 QY 9613 AATGAATATGTACAGCAACAAAAATATGAATAAT----- 9651
 Db 1612 GlnGlu-----GlnAsnProGluAsnLysValGlnGlnProLysPheCys 1626
 QY 9652 ---GAATTCAGAAAGTACTATAAAATATATATCTGCTGTTATTTGTTGAAGTGA 9708
 Db 1627 ProAspMetLysGlnProLysGlnAsnAspGlnGlnValGlyThrCysGlyLysP 1646
 QY 9709 ACACACAAAAATCATGTACTAGATGAATAATA-----AAGAAGAGAGCAACAGTT 9762
 Db 1647 GlnGlnLysLysLysValGlnLysPheValIleGlnGlnLysGlnGlnLysLys 1666
 QY 9763 CGTCTAAAGCACTCATTTCTTACACCCCATGAGATCTTCTTATACAGACCTTTA 9822
 Db 1667 AlaProGlnLysUserProProLeuThrPro-----GlnAlaProLys 1680
 QY 9823 TTCTCACACATCGAGTAGACACATATGATCT----- 9855
 Db 1681 LysGlnGlnAsnValIleProLysProProProProLysArgArgIleLysThr 1700
 QY 9856 AAAAAT-----GATATATTGAAGTACTATCTCTGTTGTTATGATATG 9900
 Db 1701 ArgAsnValIleAspHisProAlaValIleProAlaLeuMetSerSerThrIleMetTrp 1720
 QY 9901 GCGTGAAGTTTG-----ATACGCTGCTCATTTCAAGAGAAAAAATTCAAATCG 9948
 Db 1721 SerIleGlyIleGlyPheAlaAlaPheThrTyThrPheTyLeuLysLysLysThrLysSer 1740

QY 9949 TCTGTG---GACTTGTGGTACTGATATATCCCAAGAGAGATATGAAATGCTTACG 10005
 Db 1741 SerValGlyAsnLeuPheGlnIleLeuGlnIleProLysSerAspIleProThr 1760
 QY 10006 TTGCAATCCAAAAATAGTACATATACATATAGAGTGGTCCATATTAAGCAAAACATAT 10065
 Db 1761 LeuLysSerSerAsnArgIleProTyAlaSerAspArgHisLysGlyLysThrTrp 1780
 QY 10066 ATATATATGAGAGACATCTAGTGGAGATGAAGATTAATATATGTCGACCTATCTTCC 10125
 Db 1781 IleTyMetGlnLysPheSerSerGlyAspGlu---LysTyAlaPheMetSerAspThr 1799
 QY 10126 TCTGATATCTACTCATCCGAAGTGGTGAAGATTTGATATTAATATATATATGTA 10185
 Db 1800 ThrAspIleThrSerSerLysSerLysGlnGlnLysLysLysLysLysLysLys 1819
 QY 10186 CCAGTAGTCTCTTAATATTAACATGATAGAGTACTAGAACATCAAAAGGAT 10245
 Db 1820 ProGlySerProLysTyLysThrLeuIleGlnValIleGlnProSerLysArgAsp 1839
 QY 10246 ATACCAAGTAT-----GATACCAAGATTAATGATACACACGT 10284
 Db 1840 ThrGlnAsnAspIleHisAsnAspIleProSerAspIleProAsnSerAspThrProPro 1859
 QY 10285 ACGAATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10344
 Db 1860 Pro-----LleThrAspAspLysLysLysLysLysLysLysLysLysLysLys 1877
 QY 10345 TTA-----CCAATATCAGACACCAATTAATTAACAAAGTACGATATTCATATG 10395
 Db 1878 LeuGlnAsnThrGlnAsnThrGlnProAsnIleLeuHisAspAsnValAsp-----Asn 1895
 QY 10396 AATACAGAACCTAATATCTTATATCTGATATCTGATATCTGATATCTGATATCTGATATCT 10455
 Db 1896 AsnThrHisPro---ThreSerArgHisAsnMetAspGlnLysProPheIleMetSer 1914
 QY 10456 ATTCATGATAGGATTTATATCTGGAAGAAATTAATATAT----- 10497
 Db 1915 IleHisAspArgAsnLeuPheSerGlyGlnGlnLysLysLysLysLysLysLysLys 1934
 QY 10498 ---AATATTAATATGATCTAT-----ACTAATAT 10527
 Db 1935 AsnAsnProIleAsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsnAsn 1954
 QY 10538 GATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10587
 Db 1955 HisSerProTyArgAsnAspLysAsnAspLeuTyreGlyIleAspLeuIleAsnAspAla 1974
 QY 10588 CTAGTGTGCTAAACCTTATATATATGATGATGATGATGATGATGATGATGATGAT 10628
 Db 1975 Leu---SerGlyAsnHisIleAspIleTyArgGlnMetLeu 1987
 RESULT 12
 T14029
 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
 N:Alternate names: erythrocyte membrane binding protein 1 (EMPI)
 C:Species: Plasmodium falciparum
 C>Date: 20-Sep-1999 #sequence revision 20-Sep-1999 #text_change 09-Jun-2000
 C/Accession: T14029
 R:Chen, Q.; Barragan, A.; Fernandez, V.; Sundstrom, A.; Schlichterle, M.; Sahlen, A.
 J. Exp. Med. 187, 15-23, 1998
 A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PFEM
 A:Accession: T14029
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-2228 <CH>
 A:Cross-references: EMBL:AF003473; NID:92961467; PID:92961468; PIDN:AAC05730.1
 C:Genetics:
 A/Note: PCR351.2-vari
 Alignment Scores: 7.61e-78 Length: 2228
 Pred. No.:

Score:	1731.50	Matches:	771
Percent Similarity:	29.10%	Conservative:	326
Best Local Similarity:	20.45%	Mismatches:	659
Query Match:	8.90%	Indels:	2016
DB:	2	Gaps:	134

US-10-087-013-1 (1-10628) x T14029 (1-2228)

QY	97	AAAGTGCAGAAATGGTTTTGGACGGTATATGCCAAAATATTAAGACATTCATCAAAATAT	156
Db	13	glusaprlatalshtlsvalleasrpgluphecglylnlysal---hlssaprgluvalhls	31
QY	157	-----GCCAAAAGACATGTGCATTCGGTGAAGACGGATTTGGACGAAGCATTTTGGT	210
Db	32	glylualalalsasnyrvalsergltuleuylsglserleuallaserlleu	51
QY	211	GGTGGTCCCTCTACGCCAGTAAATACCATTAATTATTTATTCATATTCATGTAAATTA	270
Db	52	glylulthralsapherthvallys-----	59
Db	271	GATCATTAAGAACATTAATTTACGGTATGATGATGTG-----AAATTGAGA	318
Db	60	-----sermetglnhrgluserlystlythrgluleuileglualabanserlysalrg	77
QY	319	CATCCCTTGC-----CATGTTAGAGAACAAACCGATTGTGAGATGAAGAA	366
Db	78	asnprocslslyssasrpglylsglylasmnspalmsprghreserallysglucln	97
QY	367	TCTGATGTGGAATTAATAATACGTAATTATTAAGAAAA---AATGATGCTATACCTGT	423
Db	98	alaclytyraspanslys-----lysmetlyscysseraenglymetthrcys	113
QY	424	GGCGCCACTTAGAACACGATATGTGTGATTAATAAAGCTGGAGCTCAATATGATATAAT	483
Db	114	alaproheprhaglyleuhtlsleCysasnlyssnheproksnkselbanseransp	133
QY	484	ACCCAAATATTT---CATGATTTATTTGGGAATGTACTAGTACAGCAAAATACGAAGT	540
Db	134	serserlysalatalshtlsasrpleuvalaglvalcysmetallalatslyrgluclly	153
QY	541	GAATCATTTGTATATATCATCCA-----CATAAAGCACTTCA---GAC	582
Db	154	gluserlleuylsthrhstlyrprolystlyrasrserlystlyrproglyserasrphero	173
QY	583	GCTGTGATCGCTGTGCAGCAAGTTTGCAGATGTAGGTATTTATTCAGACGAAATGAT	642
Db	174	metcyshtmetleualatayserphelialasrplellyasrplellelvalgelylaryasp	193
QY	643	ATGTTT-----AAACCAATATGTCCATGACAAAGTA	672
Db	194	leuylrleucllyasnllyslslylsglyglahsnlglyulthrglualarglyuleu	213
QY	673	GAACGGGTCCGAGAGGTTTCAAGAAAATACATGATGTGAATGGAAGT---GAAGTA	729
Db	214	glulnlylsleuylsgluliephelelyslstlehtsaspsnleuylsasrlysgluala	233
QY	730	AAAATGATTCATTCCTGATGATCGATGTGGAATAATATTAATTAATAAGACGATGGTGG	789
Db	234	glulysarglytrahsnnglyasrpgluasrproksnpheryllyslsleualaggluasrptrip	253
QY	790	AATGTGAATACAAATTAAGATATGGGAAGCTATAACATGTGATGCATCATAT---AAATCT	846
Db	254	thralasnararglyulthrvaltrpglylalamethrcysserlysgluleuapnsenr	273
QY	847	GGATATTTTATTCATCAATCGAAGAAATATACA-----CCATATTTTCAATACCTTAA	897
Db	274	serlyrpheargalatlnrcysasasrphelcylngllyproserglnthrhlssnlyls	293
QY	898	TGC-----GGCCATAAACACGAAG-----	918
Db	294	cysahrcysasrplyssasrlysglylalamasnlaglylssrprolysalaglyasrpglyasp	313
QY	919	-----GTTCCATACCAATTTAGATTAATGTCCCTCAATATTTACGTGGTTCGACGAA	969

Dd	314	ValThrIleValProThrTrpPheAspTrpValProGlnTrpLeuArgTrpPheGln	333
Qy	970	TGGGAGAGAGATTTCGCCGAAAAAGAAATATTAAATTGAAAAAGTCAGACTCTGT	1029
Dd	334	TrpAlaGlnAspPheCysArgLysLysLysLysLysLeuGlnAsnLeuGlnLysGlnCys	353
Qy	1030	CGT---AATGACAA-----GAACCGTTATATTATGTAGCATCAATGGAACATGATGTACG	1080
Dd	354	ArgGlyLysAspLysSerAspGlnTrpArgTrpCysSerArgAsnGlyTrpAspCysGln	373
Qy	1081	ACAACATATTTGGAAAAAGGATATTTGCATTTTGGATATATACGTACTGCTTCGCAT	1140
Dd	374	GlnThrIleSerArgLysGlyLysValArgMetGlyLysGlyCysThrAspCysPhePhe	393
Qy	1141	AAATGCAAAAGTTTGGAAAGTTGGTATGGAGATCCAAACAAAGCACTTTAAAAACAAAA	1200
Dd	394	AlaCysHisSerTrpGlnAsnTrpIleAspAsnGlnArgLysGlnPheAspLysGlnLys	413
Qy	1201	GAATAATATGAAAAAGAAATACATATATTTATCGAACGATACAAACAAATTTGTCATAT	1260
Dd	414	---LysTrpThrLysGlnLysIleSerAspGlyGlyArgLysLysArgAlaValGlyLys	432
Qy	1261	ATTATAT---ACTGAATATTTATAACAAATTTTATGAAAACTTAACGAAACCGCATAT	1314
Dd	433	ThrThrLysTrpGlnGlyTrpGlnLysSerPheTrpGlnLysLeuLysAsnAspLysTr	452
Qy	1315	GCACATATAGCACTTTTAAATTTAACTTAATGAAGAAATATATTTGTAAT	1365
Dd	453	GlyThrValAspAlaPheLeuGlyLeuLeuAsnGlnLysAlaCysLysAspIleThr	472
Qy	1366	-----GAGAGA 1371	
Dd	473	AspGlyGlyLysIleAsnPheLysGlnValAsnSerGlyGlyValValGlyLysGly	492
Qy	1372	TTACACGAGAGAAAAAGATATTACTTTTACTACAGCGCGATGACAAAGCATATTAT	1431
Dd	493	SerGlyGlyThrSerGlyAlaSerGlyThrAsnAsp---GlnAsnGlyLysPheThr	511
Qy	1432	CGTTCAGAAATTGCGCAAGTGTGCCGACTGCGGGGTCAAAATGATAGTATAAATAC	1491
Dd	512	ArgSerGlnTrpCysGlnProCysProAspCysGlyValGln-----	525
Qy	1492	ACACACAAATCAGATATATGATCGTGAACGTCTAAT-----AAT 1530	
Dd	526	---HisLysGlyLysAsnGlnTrpGlnArgLysThrLysValLysLysMetArgTrpSer	544
Qy	1531	GAAGACATTAACCTTCATCGGCGTGAAGCCGACTACATATACGCTCCCTTATGAGGT	1590
Dd	545	LysLeuTrpTrpProIleAsnGlnLys-MetValIleLeuLeuLysSerLeuLysValVa	564
Qy	1591	AATGACAAAGGTGATTATACACAAAAATTA-GAAAAATTTTGT-----AACAGCTC	1640
Dd	564	LlysAspMetMetIleLeuLysLysAsnTrpLysGlnPheCysLeuThrGlnAsnSerSe	584
Qy	1641	A-----ACTATTTCAAAAGTAAAAATTAATCAAAA-- 1671	
Dd	584	AspGlySerValGlySerValThrThrGlyAlaSerGlyGlyAsnSerGlnLysLys	604
Qy	1672	-----TGGCAATGCTATTTATAAGATGAAATTAATTAATAGATGTAAACT	1715
Dd	604	SglnLeuTrpAspLysTrpLysCys---TrpLysHisAsnGlnValGlnLysValAsnVa	623
Qy	1716	GGAAACAATATCTGAATCAATATATGAT----- 1743	
Dd	623	GlnGlnGlyLysAlaGlnLysAspAspAspGlnLeuLysGlyAlaGlnGlyLeuCysIleLe	643
Qy	1744	---AATCGCTAAG----- 1752	
Dd	643	uProAsnTrpLysAsnLysGlnValSerGlnAlaLysSerGlnAsnAsnHisAlaAs	663
Qy	1753	---ATAATATCATTTCAATAATTTTGGAAATTATGCGGTACATATTTATTAAGGATAC	1808

```
Db 663 pIleGlnlysthrpHeHisaspHeTyrTyrValAlaHisMetLeuLysaspSe 683
QY 1809 TATTAAAGTGAATGACAAA---CTTAAACCTGTATTAATAT---ACAAACGACGATGTG 1862
Db 683 rIleHisrPrpGthrLysArgLeuLysSerCysIleSerSpLysThrMetLysCys 703
QY 1863 TATTGATGAATGTAACGAATTCCTTATGCTTTTGACAGATGGGTAAACAAAAGAAGA 1922
Db 703 sArgAsnGlyCysAsnLysCysAspCysPheGlnLysrPrpValLysGlnLysGluTh 723
QY 1923 AGAATGGAATAGTATTAAGAAACCTGTACAAAAGAAATATACAGCATGCTATTA 1982
Db 723 rGluThrLysrPrpLysAspHisPheLysThrGlnGluLysIleProGlnGlyTyr 743
QY 1983 TAGTAATATTAAATCTTTTGAAGGTATTTTAAAGTATGATTAACCTTGACAA 2042
Db 743 r----- 743
QY 2043 AGATGAAGCAAAATGAAAGAACTTATGAAATATATAAGAAAAAAATGACTTTC 2102
Db 744 -----PheTh 745
QY 2103 CAATTGGAAAAATATAGGACTATTATGAGAAATGCAATGAACTCTTGTAGATCACT 2162
Db 745 rThrLeuGlu-----LeuIleLeuLysLeuGlnPhe 756
QY 2163 AAAAGAACTGCCACGATATGTAAAGCAATTAATACAAAGCAAGCATGTGAACATCCA 2222
Db 756 ulysGlu-----AspThrGlnGluAsnThrGluAsnSerIle 768
QY 2223 TAAATGCAACAACAACCCGTGTGTAAACCTGTGAGGACCAACCCATAAATAT 2282
Db 768 uAspAla----- 770
QY 2283 AAAAGAAATGACCAATCTTTTAAAGAGTGCATACGAGAAAGCAAGAAATGTGTCT 2342
Db 771 -----GlnGluAla----- 773
QY 2343 TCATTAATTGAAAGAAAGGACACAGAGTATATATAACGTGGGGGTAGAGAAAGCA 2402
Db 773 ----- 773
QY 2403 CTTCAAGACAAATTTATGATATATGATTAATGATAAACAATCTAATCTGATCTGTTTC 2462
Db 773 ----- 773
QY 2463 AAATGACCATGTGATGCAAGAGGACAGAGTGTATACAAACAAGATTGTGTAGG 2522
Db 773 ----- 773
2523 AAATGATGGGAAGTGTATCGGAAACATGCGTAAAGATCAAGAAATGTTATATGCC 2582
Db 773 ----- 773
QY 2583 TCCTAGAGACACATATATGTACATCCAATTTGAAACATTTACAAACGATGATACACC 2642
Db 774 -----GlnGluLeuLysHisLeuGln----- 780
QY 2643 ACTAATGATATATTGTTGATGATTTAGTATAATTCCTTTTGGGGAGTGTCTTCT 2702
Db 780 ----- 780
QY 2703 ATCAGCAAAATATGAGCAAAACAGATATATAGATGTATTAAGAAAGAAATACCTTAA 2762
Db 781 -----LysIleLeuLysLeuGlnAsnGlu-----AsnAsnLeuAl 792
QY 2763 GGGCCCAAGAAAGTAACTGACCCAAAACCCAGACAACTATCTGTGACGATATACGTTA 2822
Db 792 aValValAsnAlaGlyThrGlu-----GlnLysThrLeu----- 803
QY 2823 CAGTTTGCAGATATAGTATATATATCGAGAAAGAAATCTCTGGAAGAAAGCGTGA 2882
Db 803 ----- 803
```

```
QY 2883 CATGCTAAAGCTGCAAGACATTTGCAAACTGTTTTGGTAAATATACATTAAGTCACTCA 2942
Db 804 -MetAspLysLeuLeuAsnHis----- 810
QY 2943 AGCCAAAGAAATGATTAATATATATGATGATGCCCCCAATATTTAAATTTGAGGAAA 3002
Db 810 ----- 810
QY 3003 TTGGTGGAGACTAATAGAGCCAAAGATATGGAAGCCATGAATGTGATATAATATT 3062
Db 811 -----GluLeuAsn-----AspAlaThrLysCysLysAspCysProle 823
QY 3063 G-----AAGGATTAATCGGGACACCAATCAACACAAAGTACTTATTCGGATATAGTGA 3116
Db 823 uProGlnGluAspSerArgLysArgLysr-----AlaHis 835
QY 3117 TCATACACCATTTGATGATTAATCCCAAAAATTAAGATGATGACCGAATGGGCAGA 3176
Db 835 pProSerPro-----AspIlePheIlePro----- 843
QY 3177 ATGCTACTCAAGCTGCACAAAAAGAGATATATAGTGAAGAGAAAGTATAGAGTGC 3236
Db 844 -----ArgProGlnGlnLysGlnAspAspGlu----- 852
QY 3237 TAAAGTAAAGATTAATGCTCAAGGCTGTACGAAGAGAGTGTACAGTGTACAGAGTG 3296
Db 852 ----- 852
QY 3297 CACAGAGCTTGTATATGATTAATATATATATATATAGATTTAGAAAGAAACAATGGAATAT 3356
Db 853 -----AsnGluAspAspAsp----- 857
QY 3357 AATATCAGATAAATCAAAAGATTAATACAGAACACCAAAATGTCTGTAGTAATAGTG 3416
Db 858 -----GlnAspGlnLysAlaArgAspAspGlnLysAlaLysGlnLysThrGlnLysSerAl 876
QY 3417 TATTAAGCTTCCACTACTGCCCCAAAATCATATACAGCAATGTATTAATTTTCTC 3476
Db 876 aThrAspThrThrThrSerLeu----- 883
QY 3477 GGAATTTATACCAACAAAATGTGTGCAAAAAGTAATTAAGTGTACTGATGAAGTGC 3536
Db 884 -----AspValCysPr 887
QY 3537 TGTCAATGTACTAACACACGATATGAAGAAATGTGAGCATATCTCCATGATACAGAAA 3596
Db 887 olleValGlyLysValLeuThr----- 894
QY 3597 TTTTGATGATTTGTGATCACAATAAGACTTTTGTATGAAAAAAGTATGATGATGAAGTAA 3656
Db 895 -----LysAspAs 897
3657 CGAAAAATATGCTTTTAGAGATTAACCAACACAGACCATGATGTGCTGTGTGTTAAAG 3716
Db 897 nGluSerLeu-----GlnAsp-----AlaCysSerLeuLysTyr 908
QY 3717 TGAATCGAAACGACAAAGGTACAGATATAAACAACAAAAAAGCGAGAAAGAAAGATAC 3776
Db 908 rGlyGlyAsnAsnSerArgLeuGlnLysrPrpArgCysValThrProSerLysGlnProThr 928
QY 3777 GGAATGTAAAAACATGATGATATATCTTTAAAGAAACAGATGAAAGAAACAAGTAAAGA 3836
Db 928 r----- 928
QY 3837 TTGTATCAAAAAAGAAATAGTATATGATATCCGATTTGCGAATGCCGAATATTAATTT 3896
Db 929 -----SerSerSpLysAsnGly----- 934
QY 3897 AGTGAAGACCCCTGTGTGTATGCCCCCTAGAAAGCAAAAGTATAGCTATCTTCT 3956
Db 935 -----AlaIleCysValProProArgArgArgLeuThrLysLysIle 950
```



```
Db 1153 ----- 1153
QY 6069 ATCAAAAGCGAGCGATCGAAAAAGCAAAAGAGCGCGCTACAAAAAGCCGGA 6128
Db 1154 ----- 1154
QY 6129 AAAAGTGAATCTTAACACAGCAAAATGCGAGCAAAACGACCGCAGCAGCACA 6188
Db 1157 nlyslleGIuasn ----- 1157
QY 6189 ACMAACAGAAAAAGACATCAACAGCAACAACAGATCTGACGTGGCACAATGT 6248
Db 1161 ----- 1161
QY 6249 AAAGCCATCTTCTGCAATTAACAGATAGCAGGGGTGAATAGAGGTTGTAATCCAA 6308
Db 1162 ----- 1162
QY 6309 AACGTATGACAAATCTCTAAATGGGTTGATTTGTAGTAAAGCTAAAGAAATGAAA 6368
Db 1174 s ----- 1174
QY 6369 TGGCATATGATGCTCTCTAGAGAAAAAATATGATAAATAATACAAATTTTAA 6428
Db 1175 ----- 1175
QY 6429 TTATGAACCTGAAAAATAAGCTGACATGATATAAAAGAGCTTTATTAATGTGAGC 6488
Db 1182 nvalIlyleuasp ----- 1182
QY 6489 AATGAAACTCAATTTTGTGTTAAATATATTAATGAAATCTCCAGCAGAAATGA 6548
Db 1187 ----- 1187
QY 6549 ATTGCAAAATGGAACATCTCCAGTGAATTAAGAAATATATATATACATATGTGA 6608
Db 1189 ----- 1189
QY 6609 TTATGAATATGTTTGTGATGATATCTAATGATAAAAAATTTAATCTGTAAC 6668
Db 1198 ----- 1198
QY 6669 AAATAGTGTACAAACCTTCTCAATGAATAAATAAAGAAAAACAGATAAAAAAGA 6728
Db 1204 ----- 1204
QY 6729 TGAAGATTACGTAAATATATTTGGAGAAAAAATTTATTTGGAGGAATGAT 6788
Db 1208 ----- 1208
QY 6789 ATATGATTAACTTATCATCTCAGACAGCAAAAGAAAAAATTAAGATATTA 6848
Db 1208 ----- 1208
QY 6849 CCAGTACATGACATGACCAAACTGACCGCTTCCCTGAAGATTTGTAAGAGCCCA 6908
Db 1209 ----- 1209
QY 6909 AATTTGACATGTTCAAGATGGCAGAGAAATTTGTATATAGAGAAAGAACACT 6968
Db 1217 rlyrPheatgtrpPhegluIutrpelygluSerPheCysarGluarGlyslYarGle 1237
QY 6969 GTTAAATTTGGAGCGGCTGTAGAAATATGAGTGTATGAT 7013
Db 1237 nlyslleGIuasn ----- 1237
QY 7014 TATAGCGTAAAGCAACAAGAA ----- 7014
Db 1235 rGlyspelygluIaIaCysaspSerIlSerThrHisaspTyrSerThrValProSerPh 1275
QY 7036 ----- 7036
Db 1275 easnCysProglCysglYlyslHisCysSerSerTylarGlysrIleGIuarglysl 1295
QY 7083 AACTGAATATGAAGACAAAGAAAAAGTTCAAAAAAGATAAAGATGCAAAAAATATA 7142
Db 1295 silegluPheHisIysGlnSerAsnIatYrGlyGln ----- 1307
QY 7143 GGATTAATCTCTCAACGTAAAGACATAGAGAAAGCAACGTGCTCATGATATTTAA 7202
Db 1308 ----- 1308
QY 7203 CARGAAATTAAGAAATTAATGTCGAATAGAGATGTGTCTGTATGCAAAAACTTCTTC 7262
Db 1318 ntrPheaspIygluPheCys ----- 1325
QY 7263 ACAACTACCAAAAAACAACACACACATCAATCATCCGATCCGATGATATGCCAGAAATC 7322
Db 1326 ----- 1326
QY 7323 GCTGATTAATGTTCTGTAAGAAATTTAACAAGTGTAGTCTCTGAAACTTTCAAAAAAGG 7382
Db 1327 rleuGIuThrItrProasp ----- 1333
QY 7383 ATCTATGATTCATACAAAAAATTAATCTGAACTTAATACCTATGAAATGTGTAGAA 7442
Db 1333 ----- 1333
QY 7443 AGCAGCATATTAATTAATCAAGAAAGCAAGAAATATATATGATATTAATCTGAAAGAAA 7502
Db 1334 -AlaAlaIySpIeGIuarg ----- 1340
QY 7503 AATTAATCTAATGAGTCTACAAAGAAAGAAAGAAAGTAATAATAGTGAATAATAA 7562
Db 1341 ----- 1341
QY 7563 TCCCTGCGATCTCTAAGAAACCTTATGACCTGATTAATATATAGGAAGAAACCTTG 7622
Db 1344 yProCysIystrAsnIySglutYrGlyIyasp ----- 1355
QY 7623 TGAATAAGAGAAAGAAATCTGTTTAAGTATGATGATGAATGTACAAATATC 7682
Db 1356 ----- 1356
QY 7683 AAAGTCTATCAGAGGAAAAAAGATATGTCTGATCTT ----- 7739
Db 1363 rlystrPhegluIhIstrGlu ----- 1382
QY 7740 CTTAAGAAATTTAGTGAATTAATAATGAAGACTTAAGATGATATATCTCTAA 7799
Db 1382 s ----- 1382
QY 7800 AATGTTCTGCAACCTGACAGAAATGAAGAAATAGACATATAAAAACTCAACTCAGA 7859
Db 1383 ----- 1383
QY 7860 GAACGGG ----- 7890
Db 1383 nasnGIyasnCysIyValSerGIyleuasnGIyasnCysaspGIlylYasaspSerIlleas 1403
QY 7891 ----- 1403
Db 1403 palalySgluIleAlalySmetIarGserSerThrItrAspVal ----- 1421
QY 7935 AACGATATGTGTACGAATGTGTGTACTTACTCCGTAGAAATTAATAATATATAGGT 7994
Db 1421 lSerasp ----- 1423
QY 7995 TTTTGAATACATATATGAAATATGAAATATAAATAAAGTGAATAAATATACACGA 8054
Db 1424 ----- 1425
QY 8055 TGACAAAGCTTT ----- 8091
Db 1425 prItrAsnThrPhegluGIyAspaspIeulYasaspIalacySgluIhIAlaAsnIlePheIy 1445
```


Db 398 LysTyrGlnAsnGlnIleSerGlyGlyArgArgGlnIleSerThrHisSer 417
QY 1264 AATAGTGA-----TATTATAACAATTTATGAAAACCTTAGAAGAACATAT 1314
Db 418 ThrGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGlyGlnGly 433
QY 1315 GCACCTAATGAC-----ACCTTTTAAATTTACTAATGAGGAAGTAT 1359
Db 434 -----AsnGlnIleLysAspValArgSerPheLeuGlnLeuSerLysGlnLysIle 451
QY 1360 TGTAAAGAGAGATACCA---GGAGAAAAGATATTACTTACTACACAGTGTGATGAC 1416
Db 452 CysLysGlnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 471
QY 1417 AAGGATATTTTATGCTTCAGATATTCGCAAGTGTGCTCCGAGTGGGATCAATGT 1476
Db 472 SerAsnThrPheSerHisThrGlnGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 491
QY 1477 GATGATATAAATACACACACCAATGAGTATGAT---CGTGAAGCTGTAAATATGAA 1533
QY 492 SerSerAspAsnCysArgLysLysProAspLysSerCysAspGlnGlnIleThrAspLys 511
Db 1534 GACTATAAACCTCCATGGGGGTGAAGCTACTAATATCAGTCTTATAGTGTAAT 1593
QY 512 GlnTyrProProGlu-----AsnThrLysIleProLysLeuThrAlaGlnLys 528
QY 1594 GAACAAGTGTATTTACACAAAAATTAGAAAATTTTGTAAACAGCTCAACATTAATCAAA 1653
Db 529 ArgLysThrGlyLeuLeuLysLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 544
QY 1654 GATTAATAATAT---CAAAATGGGATGCTATAT---AAG 1689
Db 545 AspGlyAsnAsnGlnLysGlnIleLysLysTrpGlnCysHisTrpGlnLysAsnAspLys 564
QY 1690 GATGAA-----AATATAATAGATGT---AACTGAGACAAAATAGTGAATCAAT 1737
Db 565 AspAspGlnLysAsnLysLysLysLysLysLysLysLysLysLysLysLysLysLys 584
QY 1738 AATGATATCTTAAGATATATATCATTTATTTTGAATTAAGGTTATCATATTTA 1797
Db 585 ValTyrTyrPro-----IleSerTyrTyrSerPhePheTyrGlySerIleIleAspMet 602
QY 1798 TTAAGGATATCTTAAGATATATATCATTTATTTTGAATTAAGGTTATCATATTTA 1857
Db 603 LeuAsnGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 622
QY 1858 ---CATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1914
Db 623 GlyLysCysArgLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 642
QY 1915 AAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1974
Db 643 LysLysAspLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 660
QY 1975 TCGTATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2034
Db 661 -----LeuLysAspIleAlaGlyMetAspAlaGly----- 670
QY 2035 CTGACAAAGATGACCAAAATGAAAAGTATGAAAATATATAAAGAAAAAAT 2094
Db 671 -----GlnLeuLeuGln----- 674
QY 2095 GACTTTTCAATTTGAAAAATATAGGAGCTATTTAGAGATGCAATAGAACCTGTGTTA 2154
Db 675 -----PheTyrLeuGlnLys-----IlePheLeu 682
QY 2155 GATCATTAAAGAACTGCCAGATATGTAAAGACATATATCAAAAGATGTGAA 2214
Db 683 GlnAspMetLysAsn----- 687
QY 2215 ACATCCATATATGCAACACAAACCGTGTGTAAACCTGTGAGGACGCAACCCACT 2274
Db 687 ----- 687

QY 2275 AAAAAATATAAAGAAATAGCACAAATACCTTTAAAAAGAGTGCATAGCAGGAAGCAAGAAAT 2334
Db 687 ----- 687
QY 2335 CGTGCTTCATTAATTTGAAGGAAGGACACAGGAAGTATATATAACGTGGGGGTAGG 2394
Db 687 ----- 687
QY 2395 AGAAAGACTTCAGAGCAATTTATGTAGTAATATGATAAATCTATCTATCTT 2454
Db 687 ----- 687
QY 2455 GCTTTTCAATGAGACCTGATGCGCAAGGACAGCTGATGTATACAAAGATTT 2514
Db 688 -----AlaAsnGly----- 690
QY 2515 GTGCTGAGACTGATAGGGAAGTGTGATCCGGAACACATGCGTTAAAGATCAAGAAATGTT 2574
Db 691 -----AspPro----- 692
QY 2575 ATTATGCTCTCTAGAGACACATATGTATCATCCAAATTTGAAACATTTACAAAGCGAT 2634
Db 692 ----- 692
QY 2635 GATCACCACCTTAATGTAAATTTGTTGATGATTTAGTAAATAATCTTTTGGGGGAT 2694
Db 692 ----- 692
QY 2695 GTTCTTCTATCAGCAAAATTTGAAGCAAAACAGATATAGCATGTATAAGAAAGAAAT 2754
Db 693 -----LysValIleGlnLysPheLysGln----- 700
QY 2755 AACCTAAAGGGCCCAAGAAGTAACTAGTACCACCAAAACACAGACAACTACTGTGAGCT 2814
Db 700 ----- 700
QY 2815 ATACGTTACATTTTTCAGATATATAGTATGATATTAATTCGAGGAAGATCTGGGAAAGA 2874
Db 700 ----- 700
QY 2875 AACGGTGCATGGAAGTTCGCAAGGACATTTGGAACCTGTTTTGGTATATATACATAG 2934
Db 700 ----- 700
QY 2935 TCACTCAAGCAAGCAAGAAATGATTAATATATATGATGATGATGATGATGATGATGAT 2994
Db 701 -----IleLeuGlnLysGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 712
QY 2995 AGGGAATTTGTTGGGAGGCTAATAGAGCCAAAGTATGGAGAACCATGAAATGTGATATA 3054
Db 712 ----- 712
QY 3055 AATATTTGAAGGATTAATGCGGACACACCATCAACAAAGTACTTTTCGCGATATAGT 3114
Db 713 -----LeuLysThrLys----- 717
QY 3115 GATCATCACCATGATGATGATATATCCACAAAATTAAGATGATGATGATGATGATGATGAT 3174
Db 718 -----ThrIleAspAspPheLeuGlnLysGlnLeu----- 727
QY 3175 GAATGTACTGCAAGGTGACGAAAAAGAGTATGATTAAGTGAAGGAAGTGAAGAG 3234
Db 728 -----AsnGlnAlaLysAsn----- 732
QY 3235 TGTAAAGTAAAGATATGCTCAAGGCTGTACGAAAGAGAGTGTACAGTTGTAGCAAG 3294
Db 733 CysValGlnLysAsnProAspAsnGlnCysProLysGlnLysAlaProGly----- 749
QY 3295 TGCACAGAACCTTGAATGATATATATGATATATATGATTAAGTAAGAAAGAAATGAT 3354
Db 749 ----- 749

Db	1020	AlAGlYgIyLyblys---	-ThrlleuThrglunhrlyrlnsnlyrSerlnsnValThrPheasn	1033
QY	4405	GAC-----	AAAAAGCCACCTTGGAGAGATTTGCCAAAGACCCACGTTTTTACGATGG	4455
Db	1039	GLYHlsleuThrglYThrlYrlyleuasnlglnPhealaseArpProserPheleuArgrtr	1056	
QY	4459	CTAACGGAAATGGGTACGACACTATTTGCTATTCACGACACAAAAATATTTGAAGATGGCGAC	4516	
Db	1059	MetThrGlUtrpGlYAspelnPheCySargrglUnrgrlleThrglnleuGlnlleuYlys	1078	
QY	4519	GAATAATGTAAAGTCAAATGACCACATTTGAAGTGTGATACAGAAATGTAAAGAAATGCAG	4578	
Db	1079	GLUArGcys-----	1081	
QY	4579	GACTAGCTTAATATATGTGAAAAAAGAGAGTGCATTCACACATATTAATTTACAG	4638	
Db	1081	----	1081	
QY	4639	GATGAACGGACGACAAAAAAGATTGATAGACACACATTGGTGTATGGTTACACTAT	4698	
Db	1082	----	1086	
QY	4699	ACTGGAACGATGCACAGATTTACTTGACAGAAATTTACTGCTAGTTGTTGGTATAG	4758	
Db	1087	Asngly---	1088	
QY	4759	CCTGGAAGTGCCTGTGTGTGTCAAAGAAATATACAAATTTGTTAGAAAAACAGGCTTACTAT	4818	
Db	1088	----	1088	
QY	4819	GATGCCGACAAACATTTGGGTGCACAAAATTTATTTGAAAATGACGACAAATATACTAAC	4878	
Db	1088	----	1088	
QY	4879	ATTTCGACTAAAGATTAAGTGCMAAGATTAAGTAAAGAGGACAAACAGAGTCTATTAG	4938	
Db	1088	----	1088	
QY	4939	TGGCAAAACAAAGGTCCCTAATTAACCTACAACTTAACCTGAAGAATGTGCTT	4998	
Db	1088	----	1088	
QY	4999	TTTCCTTCGTCGACTACGATATATGTTTCATGCAATGGATGGCAATTTATACAGATCCA	5058	
Db	1088	----	1088	
QY	5059	GAAGTTAAAGATGAAAATGGGTTCGCAAAAAGATTGATGAAGTGGCGACACGAAAGG	5118	
Db	1088	----	1088	
QY	5119	TACAATTTGGGTCAATACTACAAAGAAAAAAGAAAAAGAAAAATTAACCTCGAT	5178	
Db	1088	----	1088	
QY	5179	GCGCACAAATTTCTTATGAGTCCCGCTTGTAGTGTATGAAATATAGTTTTATGAT	5238	
Db	1088	----	1088	
QY	5239	TTAAGAGATATAATTTCTAGATTTGATTAATTTGGAAGATGAAAAACAAAAGCCGAGAA	5298	
Db	1088	----	1088	
QY	5299	AATTTGAAGAAAAATATTTAAACAAAAATGGAACATCAATGGCAAGAGAGATAGTACT	5358	
Db	1089	----	1093	
QY	5359	ACAGAAATCCCGGTAGTACTGCGGAAAAATTTTCTGGAACGAAAAATTAAGAAATGTGTG	5418	
Db	1093	----	1093	
QY	5419	TGGAACGCAATGATATGCGGATACAAACAGTGGTAGGATGATGGAATTAATGGAATTAAGT	5478	

QY 7636 GAAATCGTTTAAAGCTAGATTGAAATGAAATGTTACAAAAATTCAAAGTCTATCAG 7695
Db 1220 ----- 1220
QY 7696 GAGAAAAAGATATGTGTACTCCAAAGAGAACATATGTCTTAGAATTTAGAT 7755
Db 1220 ----- 1220
QY 7756 GAAATTTAAATTTGAAAGACTTAGAGATAGTAATTATCTCTMAAANAGTTCTGCAACT 7815
Db 1220 ----- 1220
QY 7816 GCACGAATGAAAGAAATAGACATATATAAAAACTTCAACTCAGAGAACGGGCGCAATG 7875
Db 1220 ----- 1220
QY 7876 AATCCAAATATGTATCTATAGAAATATAGTTTCGCTGATCGGTGACATAGTTAGAGA 7935
Db 1220 ----- 1220
QY 7936 ACAGATATGTAGCAATTTGGTGGTTACTTACCTCCGCTAGAAATTAATTAAGTT 7995
Db 1220 ----- 1220
QY 7996 TTTGAATACATATATGAGAAATGAGAAATATAAAAGTAAATTAATACACGAT 8055
Db 1220 ----- 1220
QY 8056 GTACAAACGTTTCTGCTGCTGTGGGAGCTAATAGAAAGATATTTGAAAGCAATG 8115
Db 1220 ----- 1220
QY 8116 ACGTCAAAGCACCAGAAAGATGCAAAACTTTTGAAGAAAGAAATGATGATTTGAA 8175
Db 1220 ----- 1220
QY 8176 CGCATACATTAATACAGATAGTGTGACATAGAGAGATCCACTGTGATGATAT 8235
Db 1220 ----- 1220
QY 8236 ATACCTCAACGGTTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 8295
Db 1220 ----- 1220
QY 8296 GAAGAAATGGAATAATTTAAAAATCATGTGATGATGATGATGATGATGATGATG 8355
Db 1220 ----- 1220
QY 8356 AATGATTATGAGAAATTAAGTGTGAACAGTGTAAACGAGATGTCAAGATATATAAAT 8415
Db 1220 ----- 1220
QY 8416 TTTGTTCTTAATGAAATCTCTATTCGATATACAAATCAATATAATACAAAGATTGAT 8475
Db 1220 ----- 1220
QY 8476 GAACAACAAATATATACAAAAATCTCTACTATGATGATGATGATGATGATGATG 8535
Db 1220 ----- 1220
QY 8536 TTGAAACCTTTTAAAGGATGATGTTCTGTGAGAGCTTTTCTGAATATCTCATGAAACA 8595
Db 1220 ----- 1220
QY 8596 AGTAGTGTGAAATATATATAATTAATGAAATATGATGTTCTCTCAATATACGAACATAT 8655
Db 1220 ----- 1220
QY 8656 GCTTCGAGAAACCAACAAAAAGTTATAGAGAGCTTCAGTGTGACACTACCTCTAAG 8715
Db 1221 ----- 1223
LeuProlys ---
QY 8716 AATCATTTGATTAATTTGTCTACGATCAACAAACAAAGATGATGATGATGATGATG 8775

Db 1223 ----- 1223
QY 8776 TTTACCTTCTGCTGAGAAATGATATATATATATATATATATATATATATATATAT 8835
Db 1223 ----- 1223
QY 8836 GTTCTTAATAGTTCAGATGATATACAAAGGTATTTGATTTCTCTCCAGAAAGACATTTA 8895
Db 1224 ----- 1226
ProProlys ---
QY 8896 TGTACAAAGCTATACATCTCATATATATATAGAAAAAGGTATTAAGAAATTTAAAAAA 8955
Db 1226 ----- 1226
QY 8956 AAACCTTCTACTTCT 9015
Db 1226 ----- 1226
QY 9016 GAAGAGTGTGCTTTGAGGCAATGAAATATAGTTATGCAATTTCCGATATATATAA 9075
Db 1226 ----- 1226
QY 9076 GGAACGTATATGATGACACTTCATTAATCTGAAAAATTAATAATTAATTAATCA 9135
Db 1226 ----- 1226
QY 9136 AATGAAGCAACGAAATATGTTAAAAATGATGATGATGATGATGATGATGATGATG 9195
Db 1226 ----- 1226
QY 9196 GCTATGTTATGATGATTAATAATTTCTACTCAAAAGTAACTATGATGATGATGATG 9255
Db 1226 ----- 1226
QY 9256 CAATTCAGAAAGATGAGAAAGAACTATATAGTTCTGCTGTTAATTAATGATGATG 9315
Db 1227 ----- 1229
LeuProlys ---
QY 9316 CAAGCTGTAAAGAAAGAAACATGTAAGTATTCATTTAAAAACAAATTCCTCTGTTCA 9375
Db 1229 ----- 1229
QY 9376 AACGAGTAATTTTGAAGCTCAGAAATTAATTAAGACAACTGATGATGATGATGAT 9435
Db 1230 ----- 1232
ArgInPro ---
QY 9436 AGAAATATATTAAGCTTGAATATATGATTAATAAAATACAAATGAAATCTAATATATAA 9495
Db 1232 ----- 1232
QY 9496 TATTAAGCAATTAAGAAATCAATCTTCAGTATATATAGCAATTAACATCGAAGAAAT 9555
Db 1232 ----- 1232
QY 9556 GTTACGATATATTAATAATCAAAAGATTCGAATCCGTTTGAATTAATGATTAAT 9615
Db 1233 ----- 1236
LysGluArgasp ---
QY 9616 GAAATAGTACAGAGAACAAAAATTAATGAATAATTAATCAAGAAAGTACTAAAAAA 9675
Db 1236 ----- 1236
QY 9676 TTATATCTGTTATATTTTGTGAAGATGAAACACAAAAATCATGTAATAGTGA 9735
Db 1236 ----- 1236
QY 9736 AATTAATAAGAGAGCAACAGTTCTGCTTAAAGCACTCTATTTCTTACACCCCAT 9795
Db 1237 ----- 1241
PheProProPro ---
QY 9796 GTAGATTTCTTATCAAGCACTTATATCTCAACACATGATGATGATGATGATGATG 9855
:::

Db 1242 Leu----- 1242

QY 9656 AAAAATGATATATGAAAGTAGTATCTCTGTTATATGTATCCGGCTTAGTTGATA 9915
||||| |||
Db 1243 LysAsnAlaMetLeuSerSerThrIleMetIlePheIleGlyIleGlyPheAlaThrPhe 1262
||||| |||
QY 9916 GCGCTTCATTTCAAGAGAAATAATTCAAATCGATCGGAGCTTGTCGATACGTGAT 9975
||||| |||
Db 1263 ThrIlePheIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 1282
||||| |||
QY 9976 ATCCCGCAAGAGAGTATGAAATGCTTACGTTGGATCCAAAATAGTATACCATAT 10035
||||| |||
Db 1283 IleProIleSerAspIleProIleProIleProIleProIleProIleProIlePro 1302
||||| |||
QY 10036 AGAAGTGCATATTAAGCAAAACATATATATATATGCAAGAGATACI----- 10086
||||| |||
Db 1303 ThrSerGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 1322
||||| |||
QY 10087 AGTGAGATGAGATATATATATATATGAGACTTATCTCTGATATATATCT-TCATCC 10143
||||| |||
Db 1323 SerGlyIleThrAspHisIleIle-----SerAspIleThrSerSerSer 1336
||||| |||
QY 10144 GAAAGTGAATGAAAGATTTGATATATATATATATATATATATATATATATATAT 10203
||||| |||
Db 1337 GluSerGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 1356
||||| |||
QY 10204 AAAACATTTGATAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 10257
||||| |||
Db 1357 LysThrIleIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 1376
||||| |||
QY 10258 GATACACCAACT-----AATGATACACACGATGCAATGATATTT 10296
||||| |||
Db 1377 AsnThrProSerAspThrGlnAsnAspIleGlnAsnAspGlyIleProSerSerIle 1396
||||| |||
QY 10297 ATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10356
||||| |||
Db 1397 ThrAspGlnIleThrAspThrIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 1416
||||| |||
QY 10357 GAACCAAT-----AATATATACAAAAGTCCAGATATTTCCAAATGATACAGACT 10407
||||| |||
Db 1417 GlnProAsnAspValProAsnAspIleSerSerGlyAspIleProLeuAsnThrGlnPro 1436
||||| |||
QY 10408 AATACCTTATATTTCTGATTAATCCCGAAGAAATACCTTTATATATATATATATAT 10467
||||| |||
Db 1437 AsnThrLeuIleThrAspAsnProAspIleGlyIleGlyIleGlyIleGlyIleGly 1456
||||| |||
QY 10468 GATTATATATGCGGAAAGAAATAGTATATATATATATATATATATATATATATATAT 10527
||||| |||
Db 1457 AspLeuIleSerGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 1475
||||| |||
QY 10528 GATATTCATGCAATGCGTACAGAAATGATTTCTTATAGAGTATATATATATATATAT 10587
||||| |||
Db 1476 AspIleProIleSerGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGlyIleGly 1495
||||| |||
QY 10588 CTAGTGCTGCTAAACCTTATGATATATGATGATGATGATGATGATGATGATGAT 10628
||||| |||
Db 1496 Leu--AsnSerAsnAsnValAspIleIleIleIleIleIleIleIleIleIleIle 1508
||||| |||

RESULT 14
T18399
variant-specific surface protein 1 homolog 3D7var1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18399
R:Rubio, J.P.; Thompson, J.K.; Cowman, A.F.
EMBO J. 15, 4069-4077, 1996
A:Title: The var genes of plasmodium falciparum are located in the subtelomeric region of chromosome 2
A:Reference number: 218927; MUID:96324414; PMID:8670911
A:Accession: T18399
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2042 >RUB>
A:Cross-references: EMBL:U05324; NID:91297090; PID:91297091; PIDN:AA09769.1

[illegible]

QY 1315 GCACTAATGACACTTTTAAATTAATAAGAAAGTATGTAAAGCA----- 1368
Db 308 ProaSprValAspLysPheLeuAsnLysSerLysGluThrAlaCysGlnSerGlnPro 327
QY 1369 ---GGATTACCGAGGAAAGATTAATTAATTAAGTGTGATGACAAAGGATA 1425
Db 328 TyrAspGluProAlaGlyThrIleSerIleAsnLysAsn---TyrLysAsnProAspIle 346
QY 1426 TTTATATGCTTGAATATGCGCAAGTGTCCGACGCGGGGTC----- 1470
Db 347 PheSerHisThrGluTyrCysGlnAlaCysProIlePheLysLeuPheLysAsn 366
QY 1471 ---AAATGATGATGATTAATTAATTAACACACAAATGATGATGATGATGAT 1515
Db 367 MetValAsnValGlnLysThrGlnAlaGlyAsnTyrAlaThrIleLysLeuPheLysAsn 386
QY 1516 GAACGTGTAAATTAATGAAGACTATTAACCTCAATGGGTGGAAGCTTAAATATACAT 1575
Db 387 ThrLeuIleGlnAsn-----ThrThrAspIlePro 396
QY 1576 GTCTTTATATGCTTATGAAACAGTATTAACAAATAATTAAGAAATTTTGTATC 1635
Db 397 IleLeuThrProAspThrThrLysSerAsnIleValGluLysTyrAlaGlyAsnProCysAsn 416
QY 1636 AGCTCACTAATTAACAAGATTAATAATTAATTAATTAATTAATTAATTAATTAATTAAT 1695
Db 417 SerSerAspAsnAsnSerAspGlnIleAsnAsnTyrPheLysThrAsp----- 434
QY 1696 AATATTAATTAATGAATTAATGAACAAATTAATGAATTAATGAATTAATGAATTAAT 1746
Db 435 ---GluSerLysLysSerGluLysAsnAspAsnLysValGlu 447
QY 1747 ---CCTAATATATCATTTTCAATTAATTTTGTGA 1779
Db 448 GltThrThrPheLysAsnLysAspGlnLysValThrSerTyrAsnLaphePheThr 467
QY 1780 TTATGGTTATACATATTAATTAAGGATTAATTAAGTATTAAGTATTAAGTATTAAGTATTA 1839
Db 468 LysThrValSerGluMetLeuAspSerIleLysThrAlaGlnLeuAspLysCys 487
QY 1840 AATAAATTAATTAACAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1899
Db 488 LeuLysAsnAspLysLysThrCysGluLysLysAsnTyrLysAsnLysLysAsnLysPheLe 507
QY 1900 A---GATGGTTAAACAAAGAAAGAAAGATGAATGAATGAATGAATGAATGAATGAATGA 1956
Db 507 LysAsnGluLysAsnLysLysLys-----ArgAsnGluAlaGlnLys 522
QY 1957 AAAAAGATATATCAGCAATCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2010
Db 522 LysHisThrPheLysAsnLysLysIleAspAlaGlyLysLeuValCysLeuLysLysThrHisLe 542
QY 2011 ---TATTTTAAAGTATTAAGTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2049
Db 542 LysValLeuLeuAsnAspValPheLeuGlnAspMetGluLysAlaGlnLysProGlu 562
QY 2050 ---GCAAAATGAAAGAACTTATTAAGAAATTAATAAGAAAGAAAGAAAGAAAGAAAGAAAG 2102
Db 562 HisIleLeuLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 582
QY 2103 CAATTTGGAATATTAAGGACTATTAAGAAATGAATGAATGAATGAATGAATGAATGAATGAAT 2162
Db 582 PheMetGluThrIlePheAspPhe-----LeuLeuGlnGluGlu 595
QY 2163 AAAAAGAAATGCAAGATGATTAAGAAATTAATTAACAAGAAAGAAAGAAAGAAAGAAAGAAAG 2222
Db 595 uGluGlnAspAlaGlnLysCysValSerAsnAsnPro---GluLysCysGluLysThrGlu 614
QY 2223 TAATGCAACAACAACCGCTGTGTTAACTGTGGAAGGACGACCAACCACTAAATAATAT 2282
Db 614 nLysProProThrAspGlyAla-----ProGlyGlyAla----- 625

QY 2283 AAAAGAAATGACAAATFACCTTTAAAGAGACTGCATACGAGGAAGACGAAATCGTGCT 2342
Db 625 ----- 625
QY 2343 TCATAAATTGMAAGAAAGCACACGAAAGTATATTAACGTTGGGGGTATAGAGAAAGAA 2402
Db 625 ----- 625
QY 2403 CTTCAGGACAAATTTATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2462
Db 625 ----- 625
QY 2463 AAATGACCATGATGATGCAAGGACGACGATGATGATGATGATGATGATGATGATGATGATGAT 2522
Db 626 ---GlyProSerProAspThrGlyThrAspAspAsnLeuGlu----- 638
QY 2523 AACTGAATGGAAAGTATGATCGGACACATGCGTAAAGATCAAGAAAGTATTAATGATGATGAT 2582
Db 639 ---AspIleAspSerAsp-----GlyGluGluAspAspAspVal----- 650
QY 2583 TCCTAGAAAGACATATATGTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2642
Db 651 -----SerHisValAspGluGluGluProGluAspAsnPro 662
QY 2643 ACTTAATGCTAATTAATTTGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2702
Db 662 OValGluGly----- 665
QY 2703 ATCAGCAAAATTAATGAAGCAACAAATTAATGAATTAATGAATTAATGAATTAATGAATTAAT 2762
Db 666 ---SerSerGluGluGluLysGlnGluValValLys-----AspThrGluAlaIle 681
QY 2763 GGGCCCAAGAAAGTATGATGACCCCAACACGCAACATCTGTGAGAGTATACGTTA 2822
Db 681 aValProLysGlnAspThrGlnProLysGlnGluValAsnProCysLysIleValGluGlu 701
QY 2823 CAGTTTGCAGATTAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2882
Db 701 uLeuPhe----- 703
QY 2883 CATGTTAAAGCTGCAAGGACATTTGGAACCTGTTTGGTATTAATTAATTAATTAATTAATTAAT 2942
Db 703 ----- 703
QY 2943 AGCCAAAGAAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3002
Db 704 -----LysSerThrLysAs 708
QY 3003 TTGGGGAACCTAATAGACCCAAAGATGATGGAAGCCATGAATGATTAATTAATTAATTTT 3062
Db 708 nPheGluAspAla-----CysGluLeuLysTyrGlu 718
QY 3063 GAAGGATTAATCGGACACCAATCAACACAAAGTATTATGGGATTAATGATGATCATAC 3122
Db 718 LysAsnTyrGly----- 722
QY 3123 ACCATTGATGATTAATCCCAAAAATTAAGATGATGATGATGATGATGATGATGATGATGATGAT 3182
Db 723 -----TrrLys 724
QY 3183 CTGCAAGTGCAGAAAAGAGATGATTAAGTTGAAGAGAAAGTGAAGAGATGATGATGATGATGAT 3242
Db 724 sCys----- 725
QY 3243 TAAGATTAATGTCAGAGCTGTACAAAGAGAGTGTACAGTTGTACGAAGTGCACAGA 3302
Db 725 ----- 725
QY 3303 AGCTTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3362
Db 725 ----- 725
QY 3363 AGATTAATTAACAAGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 3422

Db 726 -----ValHisHis-----ThrSerAspLysGlySerGI 735
QY 3423 AGCTTCAGTACTGCCAAAATCATATAGACAGCAATGTTATGATTTTTCGCAAT 3482
Db 735 uProthrlAlaArgGlyHisSerHisValAlaHisSer----- 747
QY 3483 ATACCAACAAATGTGGCAAAAGTAATAAAGTGTACTAGTATGAAAGTGTCTCAT 3542
Db 747 ----- 747
QY 3543 TGTACTAACACACAGTATGAAATGTGGAGCATATCTCATGATACAGAAATTTGA 3602
Db 747 ----- 747
QY 3603 TGAATGTCACTACACAAATGAGTTTGTATGAAAGTATGATGATTAACGAAA 3662
Db 747 ----- 747
QY 3663 ATATGCTTTAGATTAACCAACAGACATGATGCTGCTGTGTTAAAGTGAATC 3722
748 -----AlaAspGlyAlaProSerGlyAs 755
QY 3723 GAATCCGACAGGTACAGATTAACGAAAAAGCGGAAGAAAGATACGGAATG 3782
Db 755 pLys----- 756
QY 3783 TAAACAGTGAATGATATCTTAAGAAAAAGTGAAGAAACAGTAGAAGATTGTCA 3842
Db 756 ----- 756
QY 3843 TCCAAAAAGATAGTAATGATATCCCATGTGGCAATCGGAATATAAATTTAGTGA 3902
Db 756 ----- 756
QY 3903 AGACCTCGTGTGTATGCCCCTAGAAGACAAAGTATGCTCATTTCTGGCAAA 3962
Db 757 -AspGlyAlaIleCysIleProProlArgArgLysLeuIleHisIle----- 774
QY 3963 TGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 4022
Db 775 -----GluGlyValAspThrThrAspAspLysSerLeuArgLysTrpHeIleGlu 792
QY 4023 TGCAGACAGAGAAACATTTCTTCATGCTATTATTATA----- 4062
Db 792 rAlaAlaValGluThrPhePheLeuTrpAspArgLysLysLeuIleThrProGlnSe 812
QY 4063 -----AGTAAGATGGTGA 4076
812 rGlySerProLeuLeuGlyGlyTrpIleThrGlyValGlyValGluAsnGlyAspAspGI 832
QY 4077 AGCAATTAAGTATGATTAAGAAATTAAGAAAGCAAAATTCCTCCGCAATTTTGAGATC 4136
Db 832 uAsnAsnPro-----GluLysLeuLeuGlnLysGlyGluIleProAspGI, PheLeuArgGI 851
QY 4137 CATGTTCTACACATTTGGAGATTATAGAGATTTTATTATTGGACAGATATACAAAG 4196
Db 851 nMetPheTrpThrLeuGlyAspLysArgAspLysLeuPheSerGlyAspLysAspLysLys 871
QY 4197 TCATGT-----GAGGAGATTAACG 4217
Db 871 sAsnGlyTrpSerAspIleValSerGlyAspAsnValIleLysGluArgGluAsnThrII 891
QY 4218 AAAAGACAAATAGATTCTTTCAAAAATGCTGACAA----- 4257
Db 891 eLysGluLysIleAlaSerPhePheGlnAsnGlnLysGluGlyThrProHisValPr 911
QY 4258 -AAATCTCTAATGAAAAACGCGACAGATGTGACAGAAACATGATCATAGATATG 4316
Db 911 oLysAsnPro-----ValGlnThrProGlnThrTrpLysAspAspAlaLysHisIleTr 930
QY 4317 GGAAGCTATGCTATGTGCACTAGTAAATTTGGCGAAAAAAGATTTTACGAAAA 4376
----- 1029

Db 930 pHisGlyMetIleCysAlaLeuThr----- 938
QY 4377 CTACGCTTACAAACAGTCACAAATTTAGTGACAAAAAGCACACTTGGAGAAATTTGCCAA 4436
Db 938 ----- 938
QY 4437 ACGACCCCACTTTTACAGATGGCTAACCGAANTGTACAGACGACTATTGCTATACAGACA 4496
Db 938 ----- 938
QY 4497 AAAATATTGAAGATGTGCGAGAAAAATGTATGCAATTAACCAATGAAGTGTATAC 4556
Db 939 -----TyrGluGluLysThrSerGlySerGlyLys----- 950
QY 4557 AGAATGTAAATAGCAAAATGCGAGACTACGTTAAATATATGAAAAAAGAGTGCAT 4616
Db 950 ----- 950
QY 4617 TCCACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4676
Db 951 -----LysIleGluLysAspAspAlaValTyrLysLysPhe----- 982
QY 4677 TGTGTATAGTTACAGACTATCTGGAACGATGCAACAGATTAATTAATTAATTAATTA 4736
Db 963 -----PheGlyThrProAsnGlyAspProLeuProGlnProGI 975
QY 4737 TACTGTAGTTGCTGTATAGCAAGTCCGCTGCTGTGTACAAAGAAATATACAAAT 4796
Db 975 yThrAsnGlyThrSerAsnGluProIleSer----- 985
QY 4797 GTTAGAAAAACAGCTTACTATGATGCCGACAAACATTTGGTGCACAAAATTTATTA 4856
Db 986 -----GlnTyrGlnTyrAsp----- 990
QY 4857 AAATGACGACAAATATATACATTTTCAGATAAAGATTAAGTGAAGATTAAGTAAGTA 4916
Db 991 -----GlnValValLeuLysGI 996
QY 4917 GGCMAACAGGTCTATTAAGTGGCAAAACAAAGTCTTAAT-----AACTACAA 4967
Db 996 uGluAsnAsnGlyAlaMetSerThrSerProLysSerThrSerAlaAlaProSerAspAs 1016
QY 4968 TAACCTGAAAGAAATTAACGTAAGATGTCTTTCTTCGCTGCGACACTACATATATGTT 5027
Db 1016 nThrProThrThrLeuThrGlnPheValLeuArgProThr----- 1029
QY 5028 TCATGCATTGCATGCCAATATACAGATCCAGAAAGTTAAAGATGAAGATGGTTGCGAAA 5087
Db 1029 ----- 1029
QY 5088 AAGATTGATGGAAGTGGCGCAACGAAAGGTACAAATTTGGTCAATATACAAAGAAAA 5147
Db 1029 ----- 1029
QY 5148 AAAAGAAAAAGAAAAATTAATAACGTCGCGACCAATATTTCTTATGAGGTCCCGCC 5207
Db 1029 ----- 1029
QY 5208 TTGTAGTGTATGAATATAGTTTATATGATTTAAGAGATATATTTCTAGTTATGATA 5267
Db 1029 ----- 1029
QY 5268 TTTGGAAGATGAAGAAACGAGCAAGCAAGAAATTTGAAGAAAAATATTAACAAAAATG 5327
Db 1029 ----- 1029
QY 5328 AACATCAGTTGGCAAGGAAGTATAGTACTACAGAAATCCGGTATGACTCGCGAAA 5387
Db 1029 ----- 1029
QY 5388 ATTTTCTGGAACGAAATTAAGCAATGTGTGGAACGCAATGATATGCGGTACAAAG 5447
Db 1029 ----- 1029

```
QY 5448 TGGTAGGATGATGAAATAGTGGAAATAGTGCAGAGAGTGAAGATCTAAAAAAATG 5507
Db 1029 ----- 1029
QY 5508 TGGTCTGTACCTTCAGATGATGATATATCTATGGGAAAAATCGCATGAAGTACTGC 5567
Db 1029 ----- 1029
QY 5568 GTATCAGTTTCCTCGATGGTTTCCGCAATGGGTGAGATTTTTCGAAACATTAAGAAA 5627
Db 1030 ----- 1030
QY 5628 GGAATTTGAGAAATGTGTGGGGCTGT-----ATATATTAATCTGT--GGTGATTA 5678
Db 1047 sargtleualaginitleTyrgluaspysarglyasnaspysvalCysserglyAspCl 1067
QY 5679 TGAAGAT-----AAAAGAAA----- 5697
Db 1067 ygluaspysgluValAlcglysluAspTyserlylIeserlnpheasnCyspr 1087
QY 5698 ---TGTACAGATCGGTGTACACATATTAATAAATTTATTTAGTGAAGTGAACCCAGTA 5753
Db 1087 oglyCysglYargluCysarglyTyrlYasntrpIlelyshrlYsargthrluph 1107
QY 5754 TGAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 5813
Db 1107 easpCluInlyslYslatYrValasparglyAspLys----- 1120
QY 5814 GGCAAAAGATGCAGAGAGACGCTCGCGAATTTTAGACAAAACATTAATAAATAATTTGTA 5873
Db 1120 ----- 1120
QY 5874 AAATAAAAGTGAAGATTGTGAATATTAAGTGAAGATGTGTCCACACGCGATTAAAC 5933
Db 1120 ----- 1120
QY 5934 TGATGTATATATCAAAATATGCCCGCATTTAGACATGAACCAAAAGATGAAG 5993
Db 1120 ----- 1120
QY 5994 AAAGTAAATTTGTCAAGTGCACAGAGTCCACACAGCTGTACGAAGGAAACACCGTCACC 6053
Db 1120 ----- 1120
QY 6054 ACGGATATCATGTATATCAAAACGACGCGATCGAAAAAAGAAACGAGCGCGCC 6113
Db 1120 ----- 1120
QY 6114 TACAAAACAGCGGAAAAAGTGAATAATCTAACACAGAAATGCGACACAAACAGAAC 6173
Db 1120 ----- 1120
QY 6174 CCGACGACGACACAAACAAACAGAAACGAAATCAACAGCAACAGCAACAGATCTGA 6233
Db 1120 ----- 1120
QY 6234 CCGGCGACAAATGTAAAGCCATCTTTCGATTAACACAGATAGCAGGGGTGAATAGA 6293
Db 1120 ----- 1120
QY 6294 GGGTGTATCCAAAAGCATATGACAAATATCTAAATGGGGTGTATTTAGTAAATGC 6353
Db 1120 ----- 1120
QY 6354 TAAAGAAAATGAAATGGCATATGTATCTCTCTAGAGAAAAAATTTATGTATAATTA 6413
Db 1120 ----- 1120
QY 6414 TATACAAATTTAAATTAATGAAGTGAATGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 6473
Db 1121 ----- 1126
QY 6474 TATTAATTTGCACCAATAGCAAACTCAATTTTGTGTAAATATATATTAATGAATCC 6533
Db 1126 ----- 1126
QY 6534 TGCACAGAAATGAATTTGCAAAATGGAACAAATTCAGATGAATTTAAAGAATATGTA 6593
Db 1127 -GlyAlaGluSerLys----- 1131
QY 6594 TTATACATATGTGATTAATTAAGATATGTTTTTGGAACTGATATTTCTAATGATAAAA 6653
Db 1131 ----- 1131
QY 6654 AATTATACGTACCAAAATAGTGTAAACAACATTCATGAATAAATAAAGAAAAACA 6713
Db 1131 ----- 1131
QY 6714 GATTAATAAATAAAGATGAAGATTAAGTAAATATTTTGGGAAAAATTAATAATTTAT 6773
Db 1131 ----- 1131
QY 6774 TTGGAGAGATGATATATGATTAACCTTATCATCTACAGACGAAACGAAAAAGAAA 6833
Db 1132 ----- 1133
QY 6834 AATTAGATATATTAACAGTACATGACATGACCAAACTGACGCTTCCCTTGAAGATT 6893
Db 1133 ----- 1133
QY 6894 TGTAAAAAGCCCAATTTTGTAGATGTTTCACAGAAATGGCAGAGAAATTTGTAAATA 6953
Db 1133 ----- 1133
QY 6954 GAGAGAGAACAGTGTAAATTTGAGCGCGCTGTAGAGAAATGATGTAATGTTAG 7013
Db 1133 ----- 1133
QY 7014 TATAGCGTAAACACAGAAATGTGCAGAGCGTGTATCAATATCAAAATTTATTA 7073
Db 1134 ----- 1142
QY 7074 GAAGTGAAACTGAAATATGAAGAACACAAAGAAACTTCAAAAAGATTAAGATGCAA 7133
Db 1142 slysluenglyThr----- 1146
QY 7134 AAAGTATTAAGATTTCTCTACTGAAGAGACATGAGAGAG--GCAACATGTGCTCA 7190
Db 1147 ----- 1163
QY 7191 TGAATATTTAAAGTAAATTAATAAGATTTATGTGCATTAAGATTTCTGTATGCA 7250
Db 1163 s-----AsnAsnlyAsp----- 1167
QY 7251 AAAACCTTCTCAACTACCAAAACACACAAACATCAATCAATCGCATCTAATGA 7310
Db 1168 ----- 1171
QY 7311 TATGCCGAATTCCTGATATATGTTCTCTGAGAAATTTAAACAAGTGTGAGTGTCTGA 7370
Db 1171 ----- 1171
QY 7371 TTCAAAAAGGATCTATGATTCATACAAAAAATTAAGTAACTTAATACCTATGAA 7430
Db 1172 ----- 1178
QY 7431 TTGTGTAGAGAAAGCAGCATATTAATTTATCTAAGAGCAGAAAAATTAATGATATTA 7490
Db 1179 ----- 1181
QY 7491 CTTGAGGAAAAATTTATACATATGATGATCTACAAAGAGAAAGAAAGTAAATAATGTTG 7550
Db 1181 rPheLys----- 1183
QY 7551 GACTAATATTAATCTTGGATCTTAAGAAACCTTATGACACCTGATTAATAATATAGAGAG 7610
```

Db 1184 -----ProAlaaspasncyslys----- 1189
 QY 7611 AGAAMCCCTTGGAAATAGAGAGAAATCGTTTAAAGTAGATATGAAATGGAATG 7670
 Db 1190 -----ProCys-----SerGluPheLysVal-----LysCys 1198
 QY 7671 TTACAAAATTCAAAGTCTTATCAGAGAAAAAGAGTA---TGCTACCTCCAGAG 7727
 Db 1198 s---GluAsnGlyLysCysSerGlyGlyLysValAsnCys----- 1212
 QY 7728 AGAATATGCTTAGAATTTAGATGAATTTAAATTTGAAGACTTAGATAGTA 7787
 Db 1212 ----- 1212
 QY 7788 TTATTCCTAAATATGCTGCAACTGCAGAAATGAAGAAATACATTAATTAATA 7847
 Db 1213 -----AsnArgLysAsnThrIleAlaIleAlaThrGluIleAlaAsnMe 1226
 QY 7848 CTTCACACTCAGAGACGGGTGCGCAATGAAATCCAAATGTGATATGAAATATAGTT 7907
 Db 1226 tIleAsnSerThrGluGlyValPheMetThr---ValSerAspAsnSerAspHisLysPh 1245
 QY 7908 CGCTGATCTGGTGACATATTAGAGACAGATATGTTAGAAATTTGCTTACTTACC 7967
 Db 1245 eGlu-----GlyGlyLeuGluPr 1251
 QY 7968 TCCCGTAGAATTAATTAATAGCTTTTGAATACATATATGAAAAAGAGAAATAA 8027
 Db 1251 ocysGlySerAlaAsnIlePheLys----- 1259
 QY 8028 AATAAAGTAGAATTAATTAACACAGATGACAAACGTTTCCTGCTGGTGGATGC 8087
 Db 1260 -----G1 1260
 QY 8088 TAATAGAAAGATATTTGAAGCAATGACGTGCAAGACAGCAAGATGCAAACTTTT 8147
 Db 1260 yIleArgLysAspValTyrLys-----CysGlyGluTyrCysGlyValAspIleCys 1277
 QY 8148 TAGAAAAGAGAGATGATTTGAACGATTAACATTAATACAGATAGCTGAGACA 8207
 Db 1277 sGluProAsnThrPheAspGlyLysGluAsn----- 1287
 QY 8208 TAAAGACATCCACCTTGTATGATTAATTAATCTCAACGTTTGCATGATGATG 8267
 Db 1288 -----GlyLysGluTyrIleGluIleArg---AlaLeuLeuLysArgTr 1301
 QY 8268 GTCTGATATTATTTGAAGCAGCTGATGAGAAATTTGGAATAATTAATAATCATGGA 8327
 Db 1301 pValGluTyrPhe-----LeuGluAspTyrAsnLysIleLysHisLysIleSe 1317
 QY 8328 TCACTGTAACACATCTGACAGA-----TCCAGAAATGATGATGAAATAAGTGA 8378
 Db 1317 rHisCysMetArgLysGlyGluLysThrIleCysIleAsnAsp----- 1331
 QY 8379 TGAACAGCTGTAACAGATGTCAGAAATTAATAATTTGTTCTTAATGCAATCTCT 8438
 Db 1332 -----CysValGlu-----LysTrp----- 1336
 QY 8439 ATTGATATACAAATCAATAATAATCAAAAGATTGTATGACCAACCATATATCAAAAT 8498
 Db 1337 -IleAsnIleLysLysLysAspTrpGluThrIleArgGluArg-----TyrValLysGlu 1354
 QY 8499 CTCTACTATGATCAGTCAAAATTTTGTACAAAAGTTGAAAAGCTTTTAAAGTGAATG 8558
 Db 1354 nTyrThrThrGlyHis----- 1359
 QY 8559 TTCTGTGAGAGCTTTCTGAATATCTTCATGAACAAGTAGTGTGAATTAATTAAT 8618
 Db 1359 ----- 1359
 QY 8619 TAATGAATAATGATGTTCTTCCAAATATACGAACATATGCTTTCGAGAAAACCAAG 8678

Db 1360 -----SerAspIleTyrLysValThrSerPheLeuLysProGluInp 1374
 QY 8679 TTATTAAGAGCTTGCAGTTGTACACTACTCTTAAGAAATCCATTTGATATGCTTAC 8738
 Db 1374 eHisAsnGluValLeuLysAlaIleLysPro----- 1384
 QY 8739 CGATCAAAACAAAGATGAGTAGTAAGTAATTCAAACTTTTACCTTCTGCTCGAAGATGA 8798
 Db 1385 -----CysGlyAspLeuAspLysPheGlu----- 1392
 QY 8799 TTATGATATTAATCTGTATGATTTAGAACGATACCTGTTCTTAATAGTACAGATGA 8858
 Db 1393 -----AsnSerThrAsp----- 1396
 QY 8859 CAAAGGTATTTGATTCCTCCAGAAAGACATTTATGTACAGACCTATCATCTGCATA 8918
 Db 1397 -----CysThr----- 1398
 QY 8919 TAATTAAGAAAAGTGATTAAGAAATTTTAAAAAAAACCTTCTACTCTGCTTTCAG 8978
 Db 1399 -ValValArgSerSerGluAsnGlyValThrAsnLysLys----- 1411
 QY 8979 TCAAGACAAATTTGTAGTCAAAAATTAATTAATGGAAGAAGCTTGCTTGAGCAAT 9038
 Db 1411 ----- 1411
 QY 9039 GAAATATATGTTATGACAGATTAATCCGATTAATTAAGAAACATGATATGACACTTC 9098
 Db 1412 -----AspIleValGluCysLe 1417
 QY 9099 ATTATCTGAAAAATTTAAAAAATTTTGAACATCAAAATGACAAACGAAATCTGA 9158
 Db 1417 uLeu-----GluAsnLeuLys 1422
 QY 9159 AACATGCTGGAAAAATTAATAGAGCTCAGATATGACAGCTATGTTATGATTAATAAT 9218
 Db 1422 sThr-----LysAlaLysThrCys----- 1428
 QY 9219 TGCTACTTCAAAAGTAACATTAATGATGAAGATGCTGTCAATTAACAAAGATGAAGAAC 9278
 Db 1429 -----Pr 1429
 QY 9279 TAATCAGTTTCTGTTGTTAATTGAATGGGCAACAGCATGTAAGAAAAACA 9338
 Db 1429 oAsnGlu-----ValAsnGlyLysAsnGluThrCysAspSerLeuProH 1444
 QY 9339 TGTAGTGATTCATTAACAAACAAATGTCCTGCTCAACAGAGATTAATTTGAAGGCTC 9398
 Db 1444 sValGluAspAsp-----AspAspGluGluProLeuGluGluTh 1457
 QY 9399 AGAATTAATTAAGACACACCTGATGCTCAGAAATGATATTAAGAAATATTAAGCTTGAATAT 9458
 Db 1457 rGlu----- 1458
 QY 9459 ATTGATTAATAATCAATGCAAAATCTTAATATTAATATTAAGCAATTAAGATCATC 9518
 Db 1459 -----GluAsnThrValGluGluInProAsnIleCysProGluInLeuProLysProProPr 1476
 QY 9519 TTCAGTAATATTAACATTAATTAACATGCAATGGAAGAAAGTTTCAATATTAATTAATCAAA 9578
 Db 1476 o-----GlnProGlyAspGluAspGlyLysLysGluAlaSerProAl 1490
 QY 9579 AGATTGCAATGCGCTTGTGAGTTAATGATTAATTAATTAATTAATTAATTAATTAATTAAT 9638
 Db 1490 aProSerGlu-----GlyThrGluAs 1497
 QY 9639 TAATGAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 9698
 Db 1497 nGlu-----ProProVal----- 1501
 QY 9699 TGAAGATGAACACACAAAAAATCATGTACTGATGAATAATTAAGAAAGAGACAA-- 9756
 Db 1502 -----IleLysProGluGluGluAl 1508

```

QY 9757 -----ACAGTTCCTTAAAGACACTGATTTCTTACACCCCATGTAGATTC 9803
Db 1508 aProAlaProArgThrProArgProArgProLys-----ProProGlnGluProGln 1525
QY 9804 TTCTATCAACACCTTTATTCACACACATCGAGTAGACACATATGATCTTAAATGA 9863
Db 1525 nProTyrLeuProProAlaLeu-----LysAsnAl 1535
QY 9864 TATATGAAAAGTAGATTCCTCTGTGTATGTATGTCGCGCTTAGCTTGTGATAGCGCTCA 9923
Db 1535 aMetLeuSerSerThrIleMetTrpSerIleGlyIleGlyPheAlaThrPheTrp 1555
QY 9924 TTTCATGAAGAAAATTTCAATATGCTGTG---GACTGTGGCGTACTGATGATATCC 9980
Db 1555 eTyrLeuLysLysSerLysSerSerValGlyAsnLeuPheGlnIleLeuGlnIlePr 1575
QY 9981 GCAAGAGAGATGATGATGCTGCTGATGGAATCCAAAATAGACTTACCATATAGAAC 10040
Db 1575 oLysSerAspPheAspIleProThrLysLeuSerProAsnArgTyrIleProTyrThrSe 1595
QY 10041 TGTTCATTAATAAGGCAAAACATATATATATATGAGAGAGATACT-----AGTGG 10091
Db 1595 rGlyLysTyrGlnAlaLysArgTyrIleTyrLeuGlnGlyAspSerGlyThrAspSerGln 1615
QY 10092 AGATGAAGATAATATATGTGGGACTTATCTCTCTGATATTACT---TCATCCGAAG 10148
Db 1615 yTyrThrAspHisTyr-----SerAspIleThrSerSerSerGlnSe 1629
QY 10149 TGAGTATGAAGATTTGATATATATATATATATGACCGGTAGCCATTAATATAAAC 10208
Db 1629 rGlyTyrGlnGlnMetAspIleLeuAsnAspIleTyrValProGlyThrProLysTyrLysTh 1649
QY 10209 ATTGATAGAAGTGTACTAGACCA----- 10233
Db 1649 rLeuIleGlnValValLeuLeuProSerGlyAsnAsnThrThrAlaSerGlyAsnAsnTh 1669
QY 10234 -----TCAAAAGGATATACCAAGTGAT-----GATACACCAAGTAAATATAC 10277
Db 1669 rThrAlaSerGlyLysAsnThrProSerAspThrGlnAsnAspIleProSerGlyAspTh 1689
QY 10278 ACCACGTACGATATATATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10337
Db 1689 rPro---AsnAsnLysLeuThrAspAsnGlnTyrAsnThrLeuLysAspGlnPheIleSe 1708
QY 10338 TCATATTTTACCAATATACAGACCAAT-----AATATTTTCAAAAGCGCATAT 10388
Db 1708 rAsnMetLeuGlnSerGlnGlnProLysAspValProAsnAspTyrSerSerGlyAspIle 1728
QY 10389 TCCATGATATACAGACCAATATATATATATATATATATATATATATATATATATATATAT 10448
Db 1728 eProPheAsnThrGlnProSerThrLeuThrPheAspAsnAsnGlnGlnLysProPheIle 1748
QY 10449 TATATCATATCATGATGAGATTTATATACGGAAAGAAATAGTATATATATATATATATAT 10508
Db 1748 eThrSerIleHisAspArgAspLeuTyrThrGlyGlnGlyTyrAsnTyrAsnValAsnMe 1768
QY 10509 GAGTCTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 10568
Db 1768 tSerThrAsnSerMetAspAspIleProIleSerGlyLysAsnAspValTyrSerGlyIle 1788
QY 10569 AGATTTAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10628
Db 1788 eAspLeuLeuMetThrValCysGlyAsnHisAsnValAspIleTyrAspLeuLeu 1807
RESULT 15
T18396
eTyrLeuocyte membrane protein variant 2 - malaria parasite (Plasmodium falciparum) (fragm
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18396
R:Baruch, D.I.; Pasloske, B.L.; Singh, H.B.; Bi, X.; Ma, X.C.; Feldman, M.; Taraschi, T.
Cell 82, 77-87, 1995

```

```

A:Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen a
A:Reference number: Z18925; MID:95330812; PMID:7541722
A:Accession: T18396
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1729 <BAR>
A:Cross-references: EMBL:U27339; NID:g914920; PID:g914921; PIDN:AA08134.1
C:Genetics:
A:Gene: EMP1
A:Note: var-2

```

Alignment Scores:

```

Pred. No.: 3.91e-54 Length: 1729
Score: 1249.00 Matches: 503
Percent Similarity: 34.43% Conserved: 273
Best Local Similarity: 22.32% Mismatches: 632
Query Match: 6.42% Indels: 846
Gaps: 88

```

US-10-087-013-1 (1-10628) x T18396 (1-1729)

```

QY 97 AAAAGTCAGAAATGTTTTCAGACGTTATGCCAAAATATA----- 138
Db 13 LysAspAlaLysHisAlaLeuAspArgIleGlyGlnGluValTyrLysGlnValGlu 32
QY 139 AGACATCATCAAAATATATGCAAAAGACATGTGATTCGTTAAAGGAGATTTCAGAAA 198
Db 33 AsnAspAlaGlnLysTyrLys-----AlaLeuLysGlyAsnLeuGln 48
QY 199 GCAGATTTTCGTGTGTGTCCTTACGCCAGTAAATGACATTAATATATATATATATATAT 258
Db 49 AlalysGlyIleGlyGlnLeuAlaSer-----Proasn 60
QY 259 CCATGAAATTTGATCATATAGCAATACATCTATAT---TTACGATATGATGATGATTTG 315
Db 61 ProCysLysLeuValGlnAspTyrTyrAsnAsnArgLeuLys-----ArgLys 76
QY 316 AGACATCTCTGGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 375
Db 77 ArgTyrProCysAlaAsnArgGlnThrValArgPheSerAspGlnTyrGlyGlnCys 96
QY 376 GGA---AATTAATATCGTATATATATATATATATATATATATATATATATATATATATAT 429
Db 97 ThrPheAsnArgIleLysAspSerGlnAsnAsnAsnSerIleGlyAlaCysAlaPro 116
QY 430 CCTAGAGACGACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 489
Db 117 TyrArgGlnGlnHisLeuCysAspTyrAsnLeuGlnLysMetGlyLysThrSerThr 136
QY 480 AATATTCATGATTTATGGAATGTACTAGTACGCAAAATACGAAAGTGAATCAAT 549
Db 137 Lys---HisAspLeuLeuLeuAspValCysMetAlaIalalysTyrGlnGlyAspSerIle 155
QY 550 GTTATATATCAT---CCCATTAAGAACT-----TCAGACGCTTGACT 591
Db 156 LysThrHisTyrThrLysHisGlnLeuThrAsnProAspThrLysSerGlnLeuCysThr 175
QY 592 GCTCTGACCAAGTTTTCAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 648
Db 176 IleLeuAlaArgSerPheAlaAspIleGlyAspIleValArgGlyLysAspLeuTyrLeu 195
QY 649 -----AAACCAATGTCATGACAAAGATGAAAGGAGGATCCGAGAGGT 693
Db 196 GlyTyrAspAspLysGlnLysAspGlnArgLysLysLeuGlnAsnLeuIleGlnIle 215
QY 694 TTCAAGAAATATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 753
Db 216 PheLysLysIleHisGlnLysLeuGlnTyrGlnAspAlaLysAspHisTyrLysAsp 235
QY 754 TCTGGAATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 813
Db 236 GlnGlnLysnTyrTyrGlnLeuArgGlnAspTyrPheThrAlaAsnArgSerThrValTyr 255

```


Db 768 -----AlaThrGluGluValAspAla 774
 QY 2935 TCACGCCAAGGAAAGAAATGATAAATATATGATGATCCCCCAATATTATAATG 2994
 Db 775 AsnAlaSerSerAspGlyGluAspAspGluGluGlu----- 787
 QY 2995 AGGGAATAATGGTGGAACTAATAGACCAAAAGTATGGGAAGCATGAAATGTGTATA 3054
 Db 787 ----- 787
 QY 3055 AATATATTGAAGATATAATCGGAGACCAATCAACACAAAGTAGTTATTCGGATATAGT 3114
 Db 787 ----- 787
 QY 3115 GATCATACACATTTGATGATATATCCCAAAAATTAAGATGATACCGAATGGCA 3174
 Db 787 ----- 787
 QY 3175 GAATGTACTGCAGAGTGCAAAAAGAGTATGATAGTTGAGGAGAACTGTAAGAG 3234
 Db 788 -----Glu 788
 QY 3235 TGTAAAGATAAGGATATAGCTCAGGCTGTACGAAAGAGGTGTACAGTTGTACGAAG 3294
 Db 789 GLuGluGluGluAspGluGluGlu----- 796
 QY 3295 TGCACAAAGCTGTATGATATGATATATGATATATAGATTTATGAAAGACAATGGAAT 3354
 Db 796 ----- 796
 QY 3355 ATATATATCAGATATAATACAAAGATTAATACAGCAACCAATGTCTGTATGATAGT 3414
 Db 797 -----GluGluAlaGlu----- 800
 QY 3415 GGTATTGAAGCTTCCAGTACTGCCAAAATCATATAGACAGCAATGTATGATTTTTC 3474
 Db 800 ----- 800
 QY 3475 TCGAATTTATACCAACAAATAGTGGCAAAAGTAAATAGTGTACTAGTATGAAAGT 3534
 Db 800 ----- 800
 QY 3535 GCTGTCAATGTGTACTACACACCGTATGAAATGTTGGACATATCTCCATGATACAGA 3594
 Db 800 ----- 800
 QY 3595 AATTTGATGATTTGTAGTCACAAAATAGTTTGTGATGAAAAAGTGTGTAAGAT 3654
 Db 801 -----GluValGluGluGluGluThrAspGluSerAla 811
 QY 3655 AACGAAATAATATGCTTTAGATATAACACAGGACCATGATGTCGTGTTGTA 3714
 Db 812 ThrGluAlaValAla----- 816
 QY 3715 AGTGGATGAAACCGACAAAGGTACATATAAAAAAGAAAAAGGGAAGAAAGAT 3774
 Db 817 ---ProSerProGlyThrThrGluAspGlyValIysProAlaSerGluGluAsp 835
 QY 3775 ACGGAA---GTTAAACAGGATGATGATCTTAAAGAAAAACGATGAAAGAAACAAGTA 3831
 Db 836 ValIysValIysSerIleValAspIleAlaLeuIys-----GlyIysLeuAsp 851
 QY 3832 GAAGATTTGATCCAAAAAAGAAATAGTAATGATATCCGATTTGGCAATCGGAAATATA 3891
 Db 852 AspAlaCysThrIleuIysTyrGlyThrAlaProThrSerTrpLysCys----- 868
 QY 3892 AATTTAGTGAAGACCTCGT----- 3912
 Db 869 ---IleProSerAspThrLysSerValAlaThrThrGlySerAspThrThrGlySerGly 887
 QY 3913 ---GTGCTATGCCCCCAGAGAACAAGTTATGCTGATTTCTTG----- 3957
 Db 888 SerIleCysValProProAlaGlyArgIleuIysLeuIysValGlyLysLeuHisAspTrpAla 907

QY 3957 ----- 3957
 Db 908 GlyGlyGluThrThrGluAlaLysSerGluGluThrSerGlyGlyGluIysThrProSer 927
 QY 3958 GCMAATGATAAT-----GAAATAAAAA 3981
 Db 928 GlyAsnGluSerProSerGluLysLeuProGluGlyProThrProGluThrThrLys 947
 QY 3982 TTACATACCAAGTTAATTTAAAGAGCTTTCATCAATCTGCAGACGAGAAACATTC 4041
 Db 948 GluThrProGluSerSerLeuLeuHisAlaIleValSerProProAlaGlyLeuArgPhe 967
 QY 4042 TTCATGTGATTTATTAATAAGTAAAG----- 4068
 Db 968 LeuProThrPheLysPheLysGluGluThrPheLysAlaGluHisGlyAlaThrGly 987
 QY 4069 -----GATGTGTAAGAAATGAAATCAATCGAATTAATAA 4104
 Db 988 LeuGluLeuProGlyValThrValAspAspSerAspProAspProGluThrGluLys 1007
 QY 4105 GAAGCAAAATTCCTCCCGCATTTTGTAGATCCATGTTCTACATTTGAGATTATAGA 4164
 Db 1008 ArgGlyAsnIleProAsnAspPheLeuArgGlnMetPheTyrThrLeuGlyAspTyrArg 1027
 QY 4165 GATTTTATTTATTTGAAACAT-----ATAACAAA 4194
 Db 1028 AspIleCysIleGlyGlyAspArgAspIleValGlyAspThrIleValSerIleThrGlu 1047
 QY 4195 GGTCAATGTGAGGAAAGTAATCTAAAGCAAAAGATTTCTTTCAAAAAGTGCAC 4254
 Db 1048 GlyGluSerThrLysLysLysIleSerLysIleIleGluGlyPheLeuLysGluThr 1067
 QY 4255 CAAAAATCTCTAATGAAAA----- 4275
 Db 1068 ValThrSerProSerProArgAspThrSerArgThrProValHisProGluThrSer 1087
 QY 4276 -----ACACGCCAAGATGCTGGACAGAACATAGCATGATATGGAACT 4323
 Db 1088 ValGluLysThrProGluGluThrThrThrGluAlaAsnGlyProHisIleThrAsnGly 1107
 QY 4324 ATGCTATGTGACATAGTA----- 4341
 Db 1108 MetIleCysAlaLeuThrTyrGluLysPheSerGlyAlaIleGlyGluProGluLysVal 1127
 QY 4342 -----AAAAAT-----GGGCAAAA 4356
 Db 1128 GluAspAlaAspLysValLeuGluLysLeuLysProAsnThrAlaAsnGlyLysTyr 1147
 QY 4357 ---AAGATGATTTTTCGAAACATACGCTTACACACACGTCMAATTTAGTACAAA 4410
 Db 1148 TyrLeuLysGluAspAsnThrSerSerAlaMetProThrSerSerSerSerGly 1167
 QY 4411 AGCACC-----ACTTGAAGATTTGCCAAACGACCCAGTTTGA 4452
 Db 1168 SerAsnAspProIleAsnThrProLysLeuThrGluPheValGluIleProThrPhe 1187
 QY 4453 CGATGCTTACCGAATGTAACGACGACTTGTCTATACGCAACAAATTTGAAGAT 4512
 Db 1188 ArgTyrLeuHisGluThrPheLysGluAsnPheCysLysGluArgMetLysArgLeuGlu 1207
 QY 4513 GTGCAGGAAATAATGTAAGCAATGAC----- 4539
 Db 1208 IleTyrLysGluLysValGlyGluAsnGlyTyrGlyArgGlyArgGlyGluLysThr 1227
 QY 4540 ---CAATGAAAGT-----GAT 4554
 Db 1228 ProGluCysSerCysTyrGlyGluAspCysGluAspGluLeuSerLysTyrSerTyrAsp 1247
 QY 4555 ACA-----GAATGTAATGAATAATGCGAGACCTGTAATAAT 4593
 Db 1248 ThrValAlaAspLeuGluCysProLysCysAlaLysHisCysArgTrpTyrLysLysTrp 1267

QY 4594 ATGAAAAA-----GAGTGATTCACAGATAA----- 4629
 Db 1268 IIEGLILYSLYSASPIUPHETRGUGLGLULYSALAPHEPROLYSGILULYSASP 1287
 QY 4630 TATTACAGAGTACGACGACAAAAAGATGATAGA----- 4668
 Db 1288 TYTLYVALSNGLYASNANGLYSGLYGLYASPSANGLYPHECYSILETHLEULYS 1307
 QY 4669 -----CAACACATTTGGTTAATGCTTACACCTATACT 4701
 Db 1308 SerLeuSeraspAlaIaIaGlnPheLeuGluLysLeuGlySerCysLysAspAsnSer 1327
 QY 4702 GGAACGAAATGCAACAGATTAATCTG-----AACAGAAATTTACTGCTAGT 4746
 Db 1328 GluAspAsnGluYasnSplysAsnPheserGlnProasnGluThrPheValProAla 1347
 QY 4747 TGTGTGATTAACCTGGAAGTCCCTGTGTGTACAAAGAAATATACAAATTTAGAAAA 4806
 Db 1348 ThrAsnCysLysProCysSerGluPheLysIleAspCysLysGluAsnGlyLysCysLys 1367
 QY 4807 CAGGCTTACTATGATGCCGACAAACATTTGGGTGACAAATTTATTTGAAAAATGACGAC 4866
 Db 1368 AsnGlyLysGlyLysGlyLysGlnGluThrCysAsnGlyThrThrPheIle----- 1383
 QY 4867 AATATATCTAACATTCGAGTAAGAATAGTGAAGGATTAAGAAAGGACAAACACA 4926
 Db 1384 -----ThSerGluAsnPhesLysGlnLysGly-----GlnThrAlaLysGlu 1397
 QY 4927 GGTGCTATTAAGTGGCAAAACAAAGCTCTTAATACATTAATTAATTAATTAATTAATTAAT 4986
 Db 1398 PheValMetLysValSerAspAsnProAsnGlyPheAspAspLeuAsnGluAlaCys 1417
 QY 4987 GAAGATGTC-----CTTTTCTTCTCTGCGACTA-----CGTATATGT 5025
 Db 1418 GlnAsnAlaGlyLysPheLysSerIleArgLysAspGluThrGluCysGlyLysValCys 1437
 QY 5026 TTCATGCAATG-----GATGGCAATTAATACAGATTCAGCAAGGATAA 5067
 Db 1438 GlyTyrGluValCysIleProGluLysGlyAsnGlyValThrThrSerGlyLysAsnAsn 1457
 QY 5068 GATGAA-----AATGGTTCGCAAAAGATTAAGTGAAGTGGCGCAACGAA 5115
 Db 1458 AspGlnIleLeuThrIleArgLysLeuValAlaHisTrpValGln---AsnPhleuAsp 1476
 QY 5116 GGTACAAATTTGGGTCAATCTACAAAGAAAAAGAAAAAGAAAAATTAACACGTCG 5175
 Db 1477 AspTyrAsn-----LysIleLysHisLys 1484
 QY 5176 GATGGCAGC-----AATATCTTAAGAGTCCGCGCTTGT----- 5211
 Db 1485 ILeSerHisCysLysAsnSerSerGluGlyTyrThrCysIleLysAsnCysValGluGln 1504
 QY 5212 -----AGTGCCTAATGAATTAATTTTATGATTTAAGATTAATTTAAGTATTGAT 5265
 Db 1505 TrpIleSerThrLysArgThrGluTrpThrAsnIleLys---IleLeuLeuAsn----- 1521
 QY 5266 AATTTGCAAGATGAAAAAAGACCGAGAAAAATTTGAAGAAATA----- 5313
 Db 1522 -----GluGlnTrpLysAspAsnProAspLysAsnValLysThrIleLeuGlnAspLeu 1539
 QY 5314 -----TTTACAAAAAT-----GGAACATCATGTTGGCAAGA 5346
 Db 1540 GlnSerGlnIleAspPheAsnAlaIleLysProCysGlyThr---LeuThrLysPhe 1558
 QY 5347 AGTATAGTACTACAGAAATCCCGTAGTACTCCGCAAAATTTTCTGCAAGCAAAAT 5406
 Db 1559 GluAspSerCysGlyLysAsnGlyAlaGlnSerSerGluLysLysAsnGlyHisGlu--- 1577
 QY 5407 AAGCAATGTGTGGAACGCAATGATGCGGCTACAAACGTGTGAGC-----GAT 5457
 Db 1578 -----TyrAspAlaIleAspCysMetLeuAsnArgLeuGlnAspLysIleAsp 1593
 QY 5458 GATGAAATAGTGAATTAAGTGAAGAAAGTGAAGATCTAAAAAATGT-----GAT 5511

Db 1594 AspCysAsnLysAsnHisIleAlaGlnAsnGlyLysGluAsnGlnAlaLysCysGluLysHis 1613
 QY 5512 TCTGTACCTTCAGATGATGATTAATCTTAATGCGGAAATGCGATGAAGTACTGCGTAT 5571
 Db 1614 SerAlaProAspGluAspAspAlaIleIleGluGluGluAsnProValThrGln--- 1632
 QY 5572 CAGTTTCTTGATGATGTTGCCGAATGGGTGAGATTTTTCG---AAACATTAAGAAAG 5628
 Db 1633 -----ProAsnIleCysProLysProProGluPro 1642
 QY 5629 GAATTGCAAAATTTGTAAGGCGGTGAATGATTAATCTGTGTGTATTAATGAAGATAA 5668
 Db 1643 LysAlaGluLysGlyLysGly-----CysGluProAlaGluLysLys 1656
 QY 5689 AGAAAGAAATGTACAGATGCGGTACACAAATATTAATTAATTAATTAATTAATTAAT 5748
 Db 1657 GluLysVal-----GluGlu 1661
 QY 5749 CAGTATGAAAAACAAATCAAAAAATATGAGATTAAGCAAAATATATTCGAGCAT 5808
 Db 1662 LysGluLysThrValAsnThrValAlaLysProThrGluLysGluAlaIleGlyAsp 1681
 QY 5809 CCTGTG-----GCAAAAGATGACAGAGCGCTCGCAATATTTAGACAAACATTA 5862
 Db 1682 ProAlaGlyProAlaIleAspSerGluGluAsnProGlu----- 1694
 QY 5863 AAAATTTGTAAAAATGAAGTGGAGATTTGCAATTAATTAATTAATTAATTAATTAATTAAT 5922
 Db 1695 -----GluGluLysAlaProGluProGluValGluThrLysLysAspLysAlaPro 1711
 QY 5923 CAGGATTAATCTGATGTAATAGTCAAAATATGCCCCGATCA 5964
 Db 1712 ValLysProThr-----ProIleSer 1718

Search completed: April 28, 2003, 12:08:49
 Job time : 1240 secs

THIS PAGE BLANK (USPTO)

SUMMARIES

(without alignments)
6776.017 Million cell updates/sec

Sequence: 1 atggggttcctcgaata.....atatatgatgaagtatg 10628

Ygapop	10.0	,	Ygapext	0.5
Ygapop	10.0	,	Ygapext	0.5
Fgapop	6.0	,	Fgapext	7.0
delop	6.0	,	delext	7.0

Total number of hits satisfying chosen parameters: 1816940

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	08
1000	1000	1000

Listing first 45 summaries

Command line parameters:

```

MODEL=irameat_n2p_model -DEV-x1p
-O/-cgn2_1/USPO.spool/US10087013/rumat.28042003.102844.20247/app_query.fasta.1.10822
-DB-A-Genesecp.101002 -OEMT-fastlan -SUFFIX=n2p_rig -MINMATCH=0.1 -LOOCP=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 MATRix=blosome2 -TRANS-human40.cdl
-LIST=45 -DOCALLIG=200 -THR_SCORE=pcct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINTEN=0 -MAXLEN=2000000000
-USER=us10087013 @cgn1.1.721@umat.28042003.102844.20247 -NCP=6 -ICP=3
-NO_XLPOP -NO_MMAP -TARGETSQUARY -NEG_SCORES=0 -MAIT -LONGLOG -DEV.TIMEOUT=120
-WARR.TIMEOUT=30 -THREADS=1 -XGAP=0 -XGAPEXT=0.5 -FGAP=6 -FGAPEXT=7
-YGAP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

```

Database : A_Geneseq_101002:*

1:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1980.DAT *</i>
2:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1981.DAT *</i>
3:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1982.DAT *</i>
4:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1983.DAT *</i>
5:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1984.DAT *</i>
6:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1985.DAT *</i>
7:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1986.DAT *</i>
8:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1987.DAT *</i>
9:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1988.DAT *</i>
10:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1989.DAT *</i>
11:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1990.DAT *</i>
12:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1991.DAT *</i>
13:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1992.DAT *</i>
14:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1993.DAT *</i>
15:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1994.DAT *</i>
16:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1995.DAT *</i>
17:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1996.DAT *</i>
18:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1997.DAT *</i>
19:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1998.DAT *</i>
20:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AAI1999.DAT *</i>
21:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AI2000.DAT *</i>
22:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AI2001.DAT *</i>
23:	<i>SIDS2.gcgdata.geneseq.geneseqp_emb1/AI2002.DAT *</i>

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	19407	99.8	3542	22	AAB63142		P. falciparum FCR3
2	3501	18.0	3060	18	AAW22475		Plasmodium var-7.
3	3501	18.0	3060	21	AAV77905		Plasmodium var-7 P
4	3128.5	16.1	2913	17	AAW00384		Plasmodium falcipa
5	2972.5	15.3	2710	18	AAW22482		Plasmodium Proj3.
6	2972.5	15.3	2710	21	AAV77904		P. falciparum Proj
7	2972	15.3	2703	16	AAV70236		P. falciparum Proj
8	1773	9.1	2197	21	AAAB18352		Plasmodium falcipa
9	1755	9.0	2182	18	AAW22476		Plasmodium var-1.
10	1755	9.0	2182	20	AAV77906		Plasmodium var-1 P
11	1723.5	8.9	2228	20	AAW93944		P. falciparum PfPRM
12	1244.5	6.4	1726	15	AAW00385		Truncated Plasmodi
13	1064.5	5.5	1700	21	AAAB18144		Plasmodium falcipa
14	1032.5	5.3	700	16	AAV70235		P. falciparum EBL-
15	1032.5	5.3	700	18	AAW22481		Plasmodium ebl-2.
16	1032.5	5.3	700	21	AAV77903		P. falciparum ebl-1
17	835.5	4.3	431	21	AAAB18350		Plasmodium falcipa
18	827	4.3	407	22	AAAB63148		P. falciparum varC
19	655.5	3.4	1086	23	AAV76760		Plasmodium falcipa
20	655.5	3.4	294	22	AAAB62147		P. falciparum varC
21	649.5	3.3	921	18	AAW22480		Plasmodium E3ia.
22	649.5	3.3	921	21	AAV77902		P. falciparum ebl-1
23	636.5	3.3	1604	16	AAV70105		TNF- β -EBA 175 fusi
24	632.5	3.3	1786	14	AAAR41004		CD4-FBA175 fusion
25	631.5	3.2	793	16	AAAR70343		P. falciparum E3ia
26	618	3.2	1435	16	AAAR70332		P. falciparum SABB
27	618	3.2	1435	18	AAW22477		Silatic acid bindin
28	618	3.2	1435	21	AAV77900		P. falciparum SABB
29	609	3.1	1501	22	AAAB62150		P. falciparum varC
30	607	3.1	1501	23	AAV76762		Plasmodium falcipa
31	600.5	3.1	1421	23	AAV76764		Plasmodium falcipa
32	597	3.1	1143	23	AAV76759		Plasmodium falcipa
33	572	2.9	308	22	AAAB62151		P. falciparum varC
34	550.5	2.8	351	22	AAAB62149		P. falciparum varC
35	539.5	2.8	440	21	AAAB18146		Plasmodium falcipa
36	529	2.7	616	23	AAAM50533		Unidentified amino
37	495.5	2.5	445	22	AAAB66544		Malaria parasite
38	453.5	2.3	3973	21	AAAB18253		Plasmodium falcipa
39	431	2.2	749	16	AAAR70233		P. falciparum EBL-
40	431	2.2	749	18	AAW22479		Plasmodium ebl-1.
41	431	2.2	749	21	AAV77901		P. falciparum ebl-1
42	427	2.2	1979	21	AAAB18171		Plasmodium falcipa
43	418	2.1	10182	23	ABP38314		Staphylococcus epi
44	409	2.1	248	21	ABAB18151		Plasmodium falcipa
45	408.5	2.1	5024	22	AAAB82935		S. epidermidis ope

ALIGNMENTS

RESULT 1

ID AAB62142 standard; Protein; 3542 AA.

AC AAB62142

DT 29-MAY-2001 (first entry)
yy

P. falciparum FCR3.varCSA protein.

KM FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1,
 orthopoxvirus membrane protein 1 encoded red blood cell. PPRC

KW malaria; protozoacide
 VV

OS Plasmodium falciparum.
yy

PN WO200116326-A2
YY

PD 08-MAR-2001

XX 01-SEP-2000; 2000MO-US24195.
 PR
 XX 01-SEP-1999; 99US-0152023.
 PR
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA
 XX Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidtg C;
 PI Gysin J, Pouvelle B, Fujii N, Smith J;
 PI
 DR WPI; 2001-235109/24.
 DR N-PSDB; AAF57301.
 PT Novel FCRI3 protein, useful for modulating parasitized red blood
 PT cell binding, sequestration and onset of maternal malaria -
 PS
 XX Claim 12; Page 63-71; 78pp; English.
 CC
 CC The invention relates to a P. falciparum FCRI3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents the P. falciparum FCRI3 varCSA protein.
 CC
 SQ Sequence 3542 AA:
 Alignment Scores:
 Pred. No.: 0 Length: 3542
 Score: 19407.00 Matches: 3542
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 99.75% Indels: 0
 DB: 22 Gaps: 0
 US-10-087-013-1 (1-10628) x AAB62142 (1-3542)
 QY 1 AAGGGGTTCTCTGCAATATTTTATTAATAAAGGGAATGCACATCATCATTAAGAG 60
 DB 1 MetGlyPheSerCysLysTyrPheIleIleLysMetGlyAsnAlaLaserSerLeuGlu 20
 QY 61 GGAGAGTCATAAAGCCCTTATTAAGAAGAAAGTCAAAAGTGCAGAAAGATGTTGGAA 120
 DB 21 GlyAspAlaLysSerProIleIleLysGlySerHisLysSerAlaArgSerValLeuGlu 40
 QY 121 CGTTATGCCAAAATATTAAGACATCCATCAAAATATGCAAAAGAACATGATCGTGG 180
 DB 41 ArgTyrAlaLysAsnIleArgHisProSerLysTyrAlaLysGluHisValAspSerLeu 60
 QY 181 AAGGGGATTTACAGAAAGCAAGATTTGCTGCTCTTACGCCAGTAATAAGCAT 240
 DB 61 LysGlyAspLeuThrLysAlaGluPheArgGlyGlyProSerThrProValAsnLysHis 80
 QY 241 AATTTATATATCCATATCCATGTAATTTAGATCAATTAAGAAACATCAATTTACGGTAT 300
 DB 81 AsnTyrTyrTyrProTyrProCysAsnLeuAspHisLysGluHisThrAsnLeuArgTyr 100
 QY 301 GATGATGTGAATTTGAGACATCTTCCATCGTAGAGAAACAAACGGATTTGATAGCAT 360
 DB 101 AspAspValAsnLeuArgHisProCysHisGlyArgGluGlnAsnArgPheAspLys 120
 QY 361 GAAAGATTCGAATGTGGAATTAATTAACGTAATTTAAGAAAGAAATGATGCTATAGCC 420
 DB 121 GluGluSerGluCysGlyAsnLysLysLeuArgSerLysArgLysAsnAspAlaIleAla 140
 QY 421 TGTGGCCACCTAGAGAGACATATGTGTATAAAACCTTGAAGGCTTAATATATATA 480
 DB 141 CysAlaProProArgArgHisIleCysAspLysAsnLeuGluAlaLeuAsnAspLys 160
 QY 481 AATACCCAAATATTCATGATTTATTTGGAAATGTAAGTACTAGTACAGAAATAGCAAGCT 540

DB 161 AsnThrGlnAsnIleHisAspLeuLeuGlyAsnValLeuValThrAlaLysTyrGluGly 180
 QY 541 GATCAATTTGTTAAT 600
 DB 181 GluSerIleValAsnAsnHisProHisLysGlyThrSerAspAlaCysThrAlaLeuAla 200
 QY 601 CGAAGTTTTCAGAT 660
 DB 201 ArgSerPheAlaAspIleLysPheValArgGlyIleAspPheCysPheLysProAsnVal 220
 QY 661 CATGCAAAAGTAAAGAGGCTCCGAGAGGTTTCAGAAAGAAATATATATATATATATAT 720
 DB 221 HisAspLysValGluThrGlyLeuArgGluValPheLysLysIleHisAspLysMetGlu 240
 QY 721 GATGAAGTAAAGAAATGATATCAATCTGATGATGATGATGATGATGATGATGATGAT 780
 DB 241 AspGluValLysAsnAspLysAsnProAspGlySerGlyAsnTyrLysLeuArgGlu 260
 QY 781 GCATGTCGAATGTGAATGAATTAAGATATGGAAGCATATATATATATATATATATAT 840
 DB 261 AlaTPTPrAsnValAsnArgHisLysValIleProGluAlaIleThrCysAspAlaSerTyr 280
 QY 841 AATCTGATATTTATGCAATCAAGAAAGTAAATACCATTAATTTCAATCTTAATATGC 900
 DB 281 LysSerGlyTyrPheMetClnSerGluSerAsnThrProLeuPheSerAsnProLysCys 300
 QY 901 GGCCATTAACAAGAAAGGTTCTCAATTTAGATTTAGTCCCATATATTTAGCTGG 960
 DB 301 GlyHisLysGlnGlyLysValProThrAsnLeuAspTyrValProGluTyrLeuArgTyr 320
 QY 961 TTCAGCAATGGGGAAGAGATTTTCCGAAAGAAATATTAATTAAGAAAGGTCAG 1020
 DB 321 PheAspGluTyrGlyGluLysPheCysArgLysArgAsnIleLysLeuLysValLys 340
 QY 1021 GACCTCTGTCGTAATGCAAGAAAGCCTTAATTTGATCAATATGCAATGATTTGACG 1080
 DB 341 AspSerCysArgAsnAspLysGluArgLeuTyrCysSerHisAsnGlyHisAspCysThr 360
 QY 1081 ACNACTATTTGGAAGAAAGATTTTGCATTTGATTAATTAAGTACGATCTGCACT 1140
 DB 361 ThrThrIleTyrPheLysGlyIleLeuHisLysAsnAspLysCysThrAspCysSerThr 380
 QY 1141 AATGCAAAAGTTTGAAGTTTGGATTTAGGGAATCAACAAGACATTTAAAAACAAAA 1200
 DB 381 LysCysLysValPheGluValTyrLeuGlnLysGlnGlnGlnGlnGlnGlnGlnGln 400
 QY 1201 GAAAAATTTGAAAAAGAAATCAATCATATTTATTCAGAACGATTAACAATTTGTCAATAT 1260
 DB 401 GluLysTyrGluLysGluIleGlnSerTyrLeuSerAsnAspAsnLysPheValAsnAsn 420
 QY 1261 ATTATAGTGAATATTTAACAATTTTATGAAGAACTTAAGAAAGCAATATATGCAACT 1320
 DB 421 IleAsnSerGluTyrTyrLysGlnPheTyrGluLysLeuLysGluThrGlnTyrAlaThr 440
 QY 1321 AATGACACTTTTAAATTTAATAATTAACAAGAAAGATTTGTAAGAGATTTACAGCA 1380
 DB 441 AsnAspThrPheLeuAsnLeuAsnGlnGluLysTyrCysLysGlyLysLeuProGly 460
 QY 1381 GAAAAAGATTTACTTTTACTTAACAAGTCCATATACAAAGGATATTTTATCGTCAAA 1440
 DB 461 GluLysAspIleThrPheThrAsnSerAlaAspAspLysGlyIlePheTyrArgSerGlu 480
 QY 1441 TATTCGCAAGTGTCTCCGATGCGGGGTCAAATGATGATGATTAATTAATACACACAAA 1500
 DB 481 TyrCysGlnValCysProAspCysGlyValLysCysAspGlyIleLysTyrThrHisLys 500
 QY 1501 TCACATATATGATCTGACAGTGTAAATTAATGAAGACATTAACCTCATGGGCTGGAAG 1560
 DB 501 SerAspAsnAspArgGluArgValAsnAsnGlnLysPyrLysProProTyrGlyValLys 520
 QY 1561 CCTACTAATATCACTGCTTATATAGGTAAATGAACAAGGTATATTAACAACAAAATTA 1620
 DB 521 ProThrAsnIleThrValLeuTyrSerGlyAsnGlnGlnLysPheThrGlnLysLeu 540

QY 1621 GAAATTTTGTACAGCTCACTAATATACAAAGATAAAAATATCAAAAAATGGGAATGC 1680
|||||
Db 541 GluAsnPhcysAsnSerThrAsnTyrLysAspLysAsnAsnGlnLysTrpGluCys 560
QY 1681 TATTTATAGATGAAAAATATAAATAGATGTAACTGAGACAAAATACTGAAATCAATAT 1740
|||||
Db 561 TTTTCTTysAspGluAsnGlnLeuAsnArgCysLysLeuGlnLubGlnAsnThrGlnLeuAsnAsn 580
QY 1741 GATATCTCAAGATATATATCATTTCAATATTTTGAATATAGGGTTATACATATTTATTA 1800
|||||
Db 581 AspAsnProLysIleIleSerPheHisAsnPheGluLeuTrpValIleThrTyrLeuLeu 600
QY 1801 AGGCACTATTAAGTGAATGACAAACTTAAACTGTATTAATAATATACACACAGCAT 1860
|||||
Db 601 ArgAspThrIleLysTrpAsnAspLysLeuLysThrCysIleAsnAsnThrThrHis 620
QY 1861 TGTATGTATGAATGTAACAGAAATGTCTTATGTTTGACGATGGGTTAAACAAAAAGAA 1920
|||||
Db 621 CysIleAspGluCysAsnArgAsnCysLeuCysPheAspArgTrpValLysGlnLysGlu 640
QY 1921 GAAGAATGCAATAGTATTAAGAAGACTGTTCACAAAAAAGAAATATACAGCAATCGTAT 1980
|||||
Db 641 GluGluTrpAsnSerIleLysLysLeuPheThrLysLysAsnIleGlnGlnSerTyr 660
QY 1981 TATATGATATATTAATTAATCTTTTGAAGGTATTTTGAAGTTATGATTAACCTTGAC 2040
|||||
Db 661 TyrSerAsnIleAsnAsnLeuPheGlnGluTyrPhePheLysValMetAspLysLeuAsp 680
QY 2041 AAGATGCAACAAATGGAAGAACTTATGGAATAATTAAGAAAAAATGAGATT 2100
|||||
Db 681 LysAspGluAlaLysTrpLysGluLeuMetGluAsnIleLysArgLysLysAsnGluPhe 700
QY 2101 TCCAAATTTGAAAAATATAGGACTATTTAGAGAACTGCAATAGAACCTTGTTAGATCAC 2160
|||||
Db 701 SerAsnLeuGlnAsnAsnArgAspTyrLeuGlnAsnAlaIleGluLeuLeuAspHis 720
QY 2161 TTTAAAGAACTGCCAGATATGTAAAGACATATATACAAAGCAAGCATGTGAACATCC 2220
|||||
Db 721 LeuLysGluThrAlaThrIleCysLysAspAsnAsnThrAsnGlnAlaCysGluThrSer 740
QY 2221 CATATATCAACAAACAAACCCGTGTAAACCTCGTGAGAGCAAGCAACCCACTATAAAT 2280
|||||
Db 741 HisAsnAlaThrThrAsnProCysValLysProArgIleLysThrGlnProThrLysAsn 760
QY 2281 ATAAAGAAATAGCACAACTTTAAAGAGATGCATACAGAGCAAGCAAAATCGTGT 2340
|||||
Db 761 IleLysGluIleLeuArgIleTyrPheLysArgSerAlaTyrGluGluAlaArgAsnArgGly 780
QY 2341 CTTCATATAATTTGAAGAAAGCAACAGAAAGTATATATTAACGTGGGGGTAGAGAAAG 2400
|||||
Db 781 LeuHisLysLeuLysGlyLysAlaHisGlnGlyIleTyrLysArgGlyLysArgArgLys 800
QY 2401 GACTTCAAGGACAATTTATAGAAATATGATTAACATCTAATCTATCTGTTT 2460
|||||
Db 801 AspPheLysAspHisLeuCysArgIleMetCileLysHisSerAsnArgAsnLeuGlyPhe 820
QY 2461 TCAAAATGACCATGTGATGCAAAAGCAGAGATGATGTATACAAACAATTTGTGTA 2520
|||||
Db 821 SerAsnGlyProCysAspGlyLysGlyThrGlyAspGlyIleGlnThrArgPheValVal 840
QY 2521 GGAATGAATGGAAAGTGTATCCGGAACACATGCGTAAAGATCAAGAAATGTTATTATG 2580
|||||
Db 841 GlyThrGluTrpGlnValAlaProGlnHisMetArgLysAspHisGlnAlaSerValIleMet 860
QY 2581 CCGCTCAAGAGACAGCATATATATGATCAATTTGGAACATTTACAAAGAGATGATCAC 2640
|||||
Db 861 ProProThrArgArgHisIleCysThrSerAsnLeuGlnHisLeuGlnThrAspAspHis 880
QY 2641 CCACTTAATGATATATTTGTGATGATTTATATATATCTTTTGGGAGATGTTCTT 2700
|||||
Db 881 ProLeuAsnGlnLysAsnIleValAlaAspArgLeuValAsnAsnSerPheLeuGlyAspValLeu 900

QY 2701 CTATCAGCAAAATATGAGCAAAACAAAGATATATAGAAATGTAATAAGAAATTAACSTA 2760
|||||
Db 901 LeuSerAlaLysTyrGlnAlaAsnLysIleIleArgMetTyrLysGluLysAsnAsnLeu 920
QY 2761 AAGGGCCCCAAAGAAAGTAATCGACCCAAAACACACAACTATCTGTGACSTATACST 2820
|||||
Db 921 LysGlyProLysGluValThrAspProLysHisGlnThrThrIleCysArgAlaIleArg 940
QY 2821 TACAGTTTTCAGATATAGTGATATATATCGAGAAAGACATCTCGGAAAGAAACGGT 2880
|||||
Db 941 TyrSerPheAlaAspIleGlyAspIleIleArgGlyArgAspLeuTrpGluArgAsnGly 960
QY 2881 GACATGCTAAAGCTGCAAGACATTTGAAAATCTTTTGGTAAATATACATAAGTCACTC 2940
|||||
Db 961 AspMetValLysLeuGlnGlnLysHisLeuGluThrValPheGluLysAsnIleHisLysSerLeu 980
QY 2941 AAGGCAAGCAAGTAATATATATATGATGATGCCCCCAAAATATTTAAATGAGAGGAA 3000
|||||
Db 981 LysGlyLysGluLysAsnAspLysTyrAsnAspAspAlaProLysTyrLeuLysLeuArgGlu 1000
QY 3001 AATTGGTGGAACCTAATAGACCCAAAGTATGGGAAGCCATGAATGTATATAATAT 3060
|||||
Db 1001 AsnTrpTrpGlnAlaAsnArgAlaLysValTrpGluAlaMetLysCysAspIleLysTyr 1020
QY 3061 TTAGAGATTAATCGGACACCAATCAACAAAGTATGATTTGCGGATATAGTATCAT 3120
|||||
Db 1021 LeuLysAspLysSerGlnHisGlnSerThrClnSerSerTyrCysGlyTyrSerAspHis 1040
QY 3121 ACACCAATGGATGATATATATCCCAAAATTTAAGATGATGATGACCAATCGGAGAAATG 3180
|||||
Db 1041 ThrProLeuAspAspTyrIleProGlnLysLeuArgTrpPheThrGluTrpAlaGluTrp 1060
QY 3181 TACTGCAAGCTGCAGAAAAAGAGTATGATTAAGTTCGAAGAGAACTGTAAGAGTGAAG 3240
|||||
Db 1061 TyrCysLysValGlnLysLysGluTyrAspLysLeuLysGluLysCysLysGluCysLys 1080
QY 3241 GATTAAGATTAATGCTCAAGCTGTAGCAAAAGAGTGTACAGGTGTACGAAGTGCACA 3300
|||||
Db 1081 AspLysAspAsnGlnGlnGlyCysThrLysGlnSerCileThrGlyCysThrLysCysThr 1100
QY 3301 GAAGCTTGTAAATATATATATATATAGATTAATAGAAATTTGAAACAAACATGATATATA 3360
|||||
Db 1101 GluAlaCysAsnGlnLysTyrAsnAspIleIleGlyLeuTrpLysGlnGlnTrpAsnIleIle 1120
QY 3361 TCAGATTAATTAACAAGAAATATATATATATATATATATATATATATATATATATAT 3420
|||||
Db 1121 SerAspLysTyrLysGluLeuHisGlnGlnAlaGlnMetSerValSerAsnSerGlyIle 1140
QY 3421 GAAGCTTCCAGTCTGCCAAAAATCATATAGACAGAAATGTATATGATTTTGTGGA 3480
|||||
Db 1141 GluAlaSerSerThrAlaLysAsnHisIleAspArgAsnValIleGluPheLeuSerGlu 1160
QY 3481 TTTATACCAACAAATAGGTGCAAAAGTAAAGTAAAGTCTGTAGTATGAAAGTGCATC 3540
|||||
Db 1161 LeuTyrGlnGlnAsnGlnLysLysSerAsnLysSerLysThrSerAspGluSerAlaVal 1180
QY 3541 ATTGTACTAACAACAAGTATGAAAATGTTGAGAGCATATCTCATGATACAGAAAATTTT 3600
|||||
Db 1181 IleGlyThrAsnThrThrTyrGlnLysAsnValGlyAlaTyrLeuHisAspThrGlnAsnPhe 1200
QY 3601 GATGATTTGCACTACAAAATGAGTTTGTGATGAAAAAAGTATGATGAATTAAGTAAAGAA 3660
|||||
Db 1201 AspAspCysGlnSerGlnAsnGlnPheCysAspGluLysSerAspGlyLysAspAsnGlu 1220
QY 3661 AAATATGCTTTAGAGTAAATTAACAAGGACCAAGACATGAGTGTGCTGTATAAGTGA 3720
|||||
Db 1221 LysTyrAlaPheArgAspLysProGlnAspHisAspGlyAlaCysGlyCysLysSerGly 1240
QY 3721 TCGAAACCGCAAGGGGTACAGATAAAAACGAAAAAAGAAAGCAAGAAAGATTAAGGAA 3780
|||||
Db 1241 SerLysProThrArgValGlnIleLysThrLysLysLysArgGlnGluLysAspThrGln 1260
QY 3781 TGTAAACCACTGATGATATATCTTAAAGAAACGATGGAAGAAACAGTGAAGATTTGT 3840

```
|||||
Db 1261 CysLysThrValAsnAspIleLeuLysGluAsnAspLysLysGluValGluAspCys 1280
OY 3841 CACCACAAAAAGATATGATATGATATCCGATTTGGCAATCGCGAAATATTAATTTTACTG 3500
Db 1281 HisProLysLysAsnSerAsnGlyTyrProAspTrpIleGlySerGlyAsnIleAsnLeuVal 1300
OY 3901 GAAGACCTCGTGTGTATAGCCCTAGAGACAAAAGTTATGCGACATTTCTTGGA 3960
Db 1301 GluAspProArgValCysMetProProArgGlnLysLeuGlyValHisPheLeuAla 1320
OY 3961 AATGATATATTAATTAATAAATTTACATCACAAGTTAATTTAAAGAGCTTTCATCAA 4020
Db 1321 AsnAspAsnGlnIleLysLysLeuGlnSerGlnValAsnLeuLysGluAlaPheIleLys 1340
OY 4021 TCTGCAGCAGCAGAAACATCTCTCATGGTATTTATTAAGTAAGTAAGTGTGAAGA 4080
Db 1341 SerAlaAlaIleGlnThrPhePheSerTrpTyrTyrLysSerLysAspGlyGlnGly 1360
OY 4081 AATGAACCTCGATTAAGAAATTTAAAGAGCAAAATTCCTCCGATTTTGGATCCATG 4140
Db 1361 AsnGluLeuAspLysGlnLeuLysGlnGlyLysIleProProAlaPheLeuArgSerMet 1380
OY 4141 TCTACACATTTGGAGATTTATAGATTTTATTTTGAACAGATATATCAAAAGCTCAT 4200
Db 1381 PheTyrThrPheGlyAspTyrArgAspPheLeuPheGlyThrAspIleSerLysGlyHis 1400
OY 4201 GGTGAGGAGAACTAACTAAAGAGCAAAATAGATTTCTTTTCAAAAAGTGTGCCAAAA 4260
Db 1401 GlyGlnGlySerLysLeuLysGlnGlnIleAspSerLeuPheLysAsnGlyAspGlnLys 1420
OY 4261 TCTCTTAATGAAAAACACGCCAAGATGTGTGACAGAAACATAGTCATGATATGGAA 4320
Db 1421 SerProAsnGlyLysThrArgGlnGlnIleTrpThrGlnHisSerHisGlnIleTrpGln 1440
OY 4321 GGTATGCTATGGCACTACTATAAAATTTGGGCAAAAAAAGATGATTTTACCGAAAATC 4380
Db 1441 AlaMetLeuCysAlaLeuValLysIleGlyAlaLysLysAspAspPheThrGlnLysTrp 1460
OY 4381 GGTACACACACGCTAAATTTAGTACAAAAGCACACACTTTGGAGAAATTTGCCAAACA 4440
Db 1461 GlyTyrAsnAsnValLysPheSerAspLysSerThrThrLeuGlnGluPheAlaLysArg 1480
OY 4441 CCCGATTTTTCAGATGCTTAACCGAATGTATGACAGCACTATTTGCTATACACAGAAAA 4500
Db 1481 ProGlnPheLeuArgTrpLeuThrGlnTrpTyrAspAspTyrCysTyrThrArgGlnLys 1500
OY 4501 TATTGAAGGATGTGCAGGAAAAATGTATAGTCAAAATGACCAATTAAGTGTATACAGA 4560
Db 1501 TyrLeuLysAspValGlnGlnLysCysLysSerAsnAspGlnLeuLysCysAspThrGln 1520
OY 4561 TGTAAATTAAGAAATGCCAGAGCTAGCTTAATATATCAAAAAAAGAGTGAATTCGA 4620
Db 1521 CysAsnLysLysCysGlnAspTyrValLysTyrMetLysLysLysGlnTrpIlePro 1540
OY 4621 CAAGATTAATATTAACAGGATGACGCGCAAAAAAAGTTGATGATGACACACATTTGCT 4680
Db 1541 GlnAspLysTyrTyrLysAspIleArgAspLysLysArgPheAspArgGlnHisIleGly 1560
OY 4681 GTATAGGTTACAGATTAAGTGAACGAATGACAGATTAAGTGAACGAATTTACT 4740
Db 1561 ValMetValIleAspTyrThrGlyThrAsnAlaThrAspTyrLeuAsnArgLysPheThr 1580
OY 4741 GCTAGTTGTGATTAAGCTTGAAGTGCCTTGTGTATACAAAAGAAATATCAATTTGTTA 4800
Db 1581 AlaSerCysGlyAspLysProGlySerAlaSerValValGlnArgAsnIleGlnLeuLeu 1600
OY 4801 GAAAAACGGCTTACATGATGACCGCAAAATATTTGGGTGACAAAATTTATTAATAAT 4860
Db 1601 GlnLysGlnAlaLysTyrAspAlaAspLysHisCysGlyCysThrLysPheIleGlnAsn 1620
OY 4861 GACGACAAATATACTAAATTTCCGATTAAGTGAAGTGAAGTGAAGTGAAGGCA 4920
|||||

Db 1621 AspAspLysTyrThrAsnIleSerSerLysAspLysCysLysGlyLeuValLysGluAla 1640
OY 4921 AACACAGGTGCTATTAATAGTGGCAAAACAAAGGCTCTAATTAAGTAACTTAAGTGAAGA 4980
Db 1641 AsnThrGlyAlaIleLysThrPheGlnAsnLysGlyProAsnAsnTyrAsnAsnLeuLysGlu 1660
OY 4981 TTGACTGAAGATGTGCTTTTCTCTCTGCTGACATACGATATATGTTTTCATGATTCAT 5040
Db 1661 LeuThrGluAspValLeuPheProSerArgLeuArgIleCysPheHisAlaLeuAsp 1680
OY 5041 GGCATTTATACAGATTCACAGATTAAGATGAAGTAAAGTGGTTCGCAAAAAAGTTGATGGA 5100
Db 1681 GlyAsnTyrThrAspProGlnValLysAspGluAsnGlyLeuArgLysArgLeuMetGln 1700
OY 5101 GTGGCGGCAACGGAAGGTCATTAATTTGGTCAATCTACAAAGAAAAAAGAAAAAGAG 5160
Db 1701 ValAlaAlaIleThrGlnGlyTyrAsnLeuGlyGlnTyrTyrLysGlnLysLysGlnLys 1720
OY 5161 AAAATTAAGTGGATCGCAGCAAAATTTCTTATGAGTGGCCGCTGTGATGCTATG 5220
Db 1721 LysIleLysThrSerAspAlaHisLysTyrSerTyrGluValProProCysSerAlaMet 1740
OY 5221 AATATAGTTTATGATTTAAGAGATATATTTAGATATGATATGATATTTGGAAGATGA 5280
Db 1741 LysTyrSerPheTyrAspLeuArgAspIleIleLeuGlyIleAspAsnLeuGlnAspGln 1760
OY 5281 AATCAAAAGCCGAGAAAATTTGAAGAAAATATTTAAACAAAATGAAACATCACTGGC 5340
Db 1761 LysGlnLysThrGlnLysAsnLeuLysLysIlePheAsnLysAsnGlyThrSerValGly 1780
OY 5341 AAAGAAAGATAGTACTACAGAAATCCGGTACTACTGGCGCAAAATTTTTCGAGAC 5400
Db 1781 LysGlySerAspSerThrThrGlnLysAsnProGlySerThrAlaArgLysPhePheTrpAsn 1800
OY 5401 GAAATTAAGAAATGTGTGTGGAACGCAATGATATGCGGCTACAAACGTGTAGAGGATAT 5460
Db 1801 GlnAsnLysGlnCysValIleTrpAsnAlaMetIleCysGlyTyrLysArgGlyValArgAspAsp 1820
OY 5461 GGAATTAAGTGAAGTAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5520
Db 1821 GlyAsnSerGlyAsnSerAlaArgSerAspGlnLysPheLysLysCysGlySerValPro 1840
OY 5521 TCAGATGATGATTAATCCTATGCGGCAAAATCGCGATGAAAGTACTGCGTATACAGTTCT 5580
Db 1841 SerAspAspAspTyrPrometGlyLysAsnArgAspGlnGlyThrAlaTyrGlnPheLeu 1860
OY 5581 CGATGCTTTGCCGAATGGCGGTGAAGATTTTGCAAAACATAAAGAAAAAGAAATGGAGAA 5640
Db 1861 ArgTrpPheAlaGlnTrpGlyGluAspPheCysLysHisLysGlnLysGlnLeuGlnLys 1880
OY 5641 TTGCTAGCGCGGTGAATGATTAATTAAGTGTGCTGATTAAGTGAAGTGAAGTGAAGTGA 5700
Db 1881 LeuValGlyAlaCysAsnAspTyrThrCysGlnLysAspAsnGlnLysAspLysValGlyLysCys 1900
OY 5701 ACAGATGCTGTACACAAATATTAATAAATTTATAGTGAAGTGAAGTGAAGTGAAGTGA 5760
Db 1901 ThrAspAlaCysThrGlnTyrLysLysPheIleSerGlnTrpLysProGlnTyrGlnLys 1920
OY 5761 CAATCAAAAAATATGATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 5820
Db 1921 GlnIleLysLysTyrGlyGlnLysAsnLysAspLysIleTyrSerGlnHisProValAlaLys 1940
OY 5821 GATCAGACAGAGCTCGCAATATTTTGAACAAACAATTAATAAATAAATTTGGAATAATA 5880
Db 1941 AspAlaGlnLysAlaArgGlnTrpLeuAspLysGlnLeuLysLysIleCysGlnLysAsnLys 1960
OY 5881 AGTGAAGATGTGGAATATTAAGTGTATGAAGATGTGTCCACACAGCATTAAGTATGATGT 5940
Db 1961 SerGlnLysCysGlnTyrLysCysMetLysAspValSerThrGlnArgLeuThrAspGly 1980
OY 5941 AATAGTCAAAATATGCGCGCATCTTATGACGATGACCAAAAGAAAGTGAAGAAAGTGT 6000
Db 1981 AsnSerGlnAsnMetProValSerLeuAspAspIleProLysGlnValGlnGlyLysCys 2000
|||||
```


QY	6001	AATGTCAAGTGCACAGAGGTCCACCAAGTGTACGAGAGGAAACCGTCAACAGGGTA	6060
Db	2001	AsnGysGlnValProArgGlyProProArgValArgArgGluThrProSerProArgVal	2020
QY	6061	TCACGATATCAAAAGCGAGCGATCGAAAAAAAGAAAGCAAAACGCCCGCTACAAA	6120
Db	2021	SerLeuIleSerIysAlaThrAlaSerIysLysGluAlaIysThrAlaProPheThrLys	2040
QY	6121	CAGCCGAAAAAGTGGAAATCTACACAGAAATGCGAGCAAAACAGAACCGACGA	6180
Db	2041	GlnProLysLysValGluAsnLeuThrThrGluMetArgAlaGlnThrArgThrArgArg	2060
QY	6181	GCACACACAAACACGAAACGAAACGAATACACAGCAACAACAGATCTGACGTGGC	6240
Db	2061	AlaAlaGlnGlnThrArgLysArgThrSerThrAlaThrThrThrGluSerAspValGly	2080
QY	6241	ACAAATGTAAAGGCCATCTCTTTCGATTAACACAGATAGCAGGGCGTAATAGAGGTGT	6300
Db	2081	ThrMetValLysAlaIleLeuSerAsnLysProAspSerArgGlyGlyIleGluGlyCys	2100
	6301	AATCCAAAAACGATGGACAAATTCCTAAATGGGGTTGATTGTAGTAAGTCTAAAGA	6360
		AsnProLysThrTyrGlyGlnTyrProLysTrpGlyCysIleValGlyLysSerLysGlu	2120
	2101	AsnProLysThrTyrGlyGlnTyrProLysTrpGlyCysIleValGlyLysSerLysGlu	2120
QY	6361	AATGAAATGGCATTATCTATTCGCTCCCTACGAGAAAAAATTAATGTATTAATATACAA	6420
Db	2121	AsnGlnAsnGlyIleCysMetProProArgArgLysLysLeuCysIleAsnAsnIleGln	2140
QY	6421	TATTTAAATATATGAACCTGAAATTAAGCGCTGCACATGATATATAAGAGCGCTTTATTTAA	6480
Db	2141	TyrLeuAsnTyrGluThrGluAsnLysArgAspAsnAspIleLysGluAlaPheIleLys	2160
QY	6481	TGTGCACCATATGAACCTCAATTTTGTGTGTTAAATATATATTTGAAATCCCTGCAGCA	6540
Db	2161	CysAlaAlaIleGluThrGlnPheLeuThrPheLysTyrIleGluAsnProAlaIa	2180
QY	6541	GAAATGAATTCGAAATGGAACAAATTCACAGATGATTAATTAAGATATATATATACA	6600
Db	2181	GluAsnGlnLeuGlnAsnGlyThrIleProAspGluPheLysArgIleMetTyrTyrThr	2200
QY	6601	TATGCGTATATATAAGATATGTTTTTTTGGACACTGATATTTCTAATGATAAAAAATATTA	6660
Db	2201	TyrGlyAspTyrLysAspMetPhePheGlyThrAspIleSerAsnAspLysLysIleIle	2220
QY	6661	ACTGTACCAATAGTGTACACAAACCATTCATCAATGAATAATATACAAAAACAGATCAA	6720
Db	2221	ThrValThrAsnSerValThrThrIleLeuAsnGluAsnAsnLysLysLysGluAspLys	2240
	6721	AAAAAGATGAAGATTCGTAATAAATATTTTGGCAAAAAATATAAAATTTATTTGGCA	6780
	2241	LysLysAspGlnGluLeuArgLysIlePheTrpGluLysAsnLysLysPheIleTrpGlu	2260
	6781	GGATGATATATAGGATTAACCTTATCCATCCACACAGCAACGAAACGAAAAAATATAGA	6840
Db	2261	GlyMetIleTyrGlyLeuTyrThrIshSerThrAspGluAsnGluLysGluLysIleArg	2280
QY	6841	GATATATACAGTACATGACATGACCAAACTGACGCTCTCCCTGGAAGAGTTGTAAA	6900
Db	2281	AspAsnTyrGlnTyrAsnAspMetThrLysLeuThrProSerLeuGlnLysPheValLys	2300
QY	6901	AGGCCCAATTTTGGAGATGGTTCACACGAATGGGCAAGAAATTTTGTATATACAGCAAG	6960
Db	2301	ArgProGlnPheLeuArgTrpPheThrGlnTrpAlaGlnGluPheCysAsnLysArgLys	2320
QY	6961	GAAACAGTTGTTAAAAATGGAGCGCGGCTGTAAAGAAATATGACTGTAAATGATAGTATAC	7020
Db	2321	GluGlnIleuLeuLysLeuGluAlaGlyCysLysGluTyrGluCysAsnGlySerAsnAsp	2340
QY	7021	GGTAAACACAGAAATGTGCAGAGCGGTGTACATATCAAAATTTTATTAAGAAGTGG	7080
Db	2341	GlyLysThrGlnGlnCysAlaGluAlaCysValThrTyrGlnAsnPheIleLysLysTrp	2360

QY	7081	AAACTGATATATGAAGACCAAGAGAAAGCTTCAAAAAGGATAAAGATGCCAAAAAGTAT	7140
Db	2361	LysThrGluTyrGluAArgGlnAArgGluLysPheLysLysAspLysAspGluLysLysTyr	2380
QY	7141	AAGGATTAATCCTTCTACTGAAAGGACATATGACAGGCAACATGTCATCAATGATATTTA	7200
Db	2381	LysAspTyrProSerThrGluAArgAspLLeGluLysAlaThrCysAlaHisGluLysTyrLeu	2400
QY	7201	AACATGAAATTAAGAAATTAATGATGCGCAATATAGGATGTTCCTGTATGCAAAAACCTCT	7260
Db	2401	AsnMetLysLeuLysGluLeuLysGluLysAsnLysAspCysSerCysMetGlnLysProSer	2420
QY	7261	TCACACTACCAAAAACACACACAAATCAATCATCCGATGCTAATGATATATGCCAGAA	7320
Db	2421	SerGlnLeuProLysThrThrGlnGlnSerGlnSerSerAspAlaAsnAspMetProGlu	2440
QY	7321	TCGCTGGATTAATGTCTCTGAAAGAAATTTACCAAGTGGAGTGTCCCTGAACTTCAAAAAG	7380
Db	2441	SerLeuAspTyrValProGluGlnPheAsnLysCysGluCysProGluLeuSerLysLys	2460
QY	7381	GGATCTATGATTTCAATACAAAAAAATATACGACCTTAATAATCCTATGAATGTCTAAG	7440
Db	2461	GlySerMetLLeHisThrLysLysLLeThrGluProLysLLeProMetLAsnCysValGlu	2480
QY	7441	AAAGCAGCATATTAATTAATCTAAAGAACGAGAAATATATGGAATTAATCTTGAAGAA	7500
Db	2481	LysAlaAlaTyrTyrLeuSerLysGluAlaGluAsnAsnMetAspLLeThrLeuLysGlu	2500
QY	7501	AAATTTATACCTATTTAGTGTCTACAAAGGAAAGGAAAGTAAATATAGTTGGACTAATAT	7560
Db	2501	LysPheLLeProLLeGluSerThrLysGluLysGluSerLysAsnSerThrPheAsnAsn	2520
QY	7561	AATCCTTCGATCTCTAAGAAACCTTATGACCTGATTAATATATATGGAAGAAAGAAACCT	7620
Db	2521	AsnProCysAspProLysLysProTyrAlaProAspLysTyrLLeGluYArgTArgAsnPro	2540
QY	7621	TGTGAAATATAGACAGAAATTCGTTTAAAGCTAGATTAATGAAATGTTCACAAAT	7680
Db	2541	CysGluAsnAArgGluGlnAsnAArgPheLysValAspTyrGluTrrpLysCysTrrLysAsn	2560
QY	7681	TCAAGTCTCTACGAGGAGAAAAAGAGTATGTCTACCTCCAAAGAGAGACATATGTGC	7740
Db	2561	SerLysPheTrrGlnGluLysLysAArgValCysValProProArgTArgLLeHisMetCys	2580
QY	7741	TTAAGGATTTAGATGGAATTTAAATTTGAAGACCTTAAGAGATAGTAAATATATCTCTAAA	7800
Db	2581	LeuArgAsnLeuAspGluLLeLysLLeGluLArgLeuLysAspSerAsnTrrLeuLeuLys	2600
QY	7801	ATGTTCTGTCACACCTGCACAGAAATGAAGCAATAGACATATATATAAAACCTTCAACTCAG	7860
Db	2601	MetValAArgTrrAlaAArgAsnGluGluLLeAspLLeLLeLysAsnPheAsnSerLLeu	2620
QY	7861	AACGGGTCCGCAATGAATCCAATATGTGATCTATGAATATATAGTTTGCCTGATCTGGGT	7920
Db	2621	AsnGluCysAlaMetCAsnProLLeCysAspThrMetLysTrrSerPheAlaAspLeuGly	2640
QY	7921	GACATAGTTTAAGGACACAGATATGTTACGAATGTGGTGTACTTACCTCCCTGCAAAATA	7980
Db	2641	AspLLeValAArgGluTrrAspMetLeuArgLLeGluTrrLeuProProValGluLLe	2660
QY	7981	AAATTAATTAAGCTTTTGAATACATATATGAGAAATGAGAAATATAAATTAAGCTGACA	8040
Db	2661	LysLeuTrrLysValAPheGluTrrLLeTrrGlyLysTrrPArgAsnLysAsnLysGluArg	2680
QY	8041	AATAAATACACAGATGTACAAACGTTTCTGTCTGTGGTGGAGTCTATATGAAAAAGAT	8100
Db	2681	AsnLysTrrAsnAspValGlnThrPheArgSerAlaTrrPrrAspAlaAsnArgLysAsp	2700
QY	8101	ATTGGAAAGCAATGACGTCGAAAGCAACCGAAGATGACAAACCTTTTGAAGAAAGAGAGA	8160
Db	2701	LLeTrrpLysAlaMetTrrCysLysAlaProGluAspAlaLysLeuPheArgLysGluArg	2720
QY	8161	ATGCATGATTTGAACCATATACATTAATCAAGATATAGTGGACATATAGACATATCA	8220

QY 10381 GCAGATATTCACATGAATACAGACCTAATACCTTATATTCGATATATCCTGAAGAAA 10440
 |||||
 Db 3461 AlasplleprokelastrnrlgluproasnrlhLeuTyseraspnsprogluLulys 3480
 QY 10441 CCTTTATATATATTCATTCATGATAGGATTTATATCTGGAGAAATTTACTATAT 10500
 |||||
 Db 3481 ProhellellelesterlelehisaspargaspLeuTythrGlyserGluileseTyasn 3500
 QY 10501 ATTAATATGCTACTATATATATATGATATCCATGATGCTGAATGTTCTAT 10560
 |||||
 Db 3501 IleasmetserThrasnrlhAsnaspIleprokelastrnAlaArgasnaSPserTyr 3520
 QY 10561 AGAGTATATATTAATTAATGATTCACATGCTGCTAAACCTATGATATATGATG 10620
 |||||
 Db 3521 ArgGlyIleAspleuleleAsnaspserleuValValleuAsnleuleuIeTyrmelMe 3540
 QY 10621 AAGTAT 10626
 |||||
 Db 3541 LysTyr 3542
 |||||
 ID AAM22475 standard; Protein; 3060 AA.
 AAM22475
 AC AAM22475;
 XX
 DT 12-SEP-1997 (first entry)
 XX
 DE Plasmodium var-7.
 XX
 KW DBL gene family; SABP, stalloic acid binding protein; vaccine; therapy;
 KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KM Plasmodium.
 XX
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 XX
 PN MO9640766-A2.
 XX
 PD 19-DEC-1996.
 XX
 PF 07-JUN-1996; 96WO-US09508.
 XX
 PR 07-JUN-1995; 95US-0487826.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 Wellens TE;
 WPI: 1997-052231/05.
 N-PSDB; AAT72882.
 DR
 DR
 XX
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stalloic acid
 PT binding proteins
 PS
 PS Claim 8: Page 61-67; 96p: English.
 XX
 CC This sequence represents var-7 of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stalloic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of

CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).
 CC
 SQ Sequence 3060 AA;
 Alignment Scores:
 Pred. No.: 5,39e-238 Length: 3060
 Score: 3501.00 Matches: 1057
 Percent Similarity: 38.65% Conservative: 470
 Best Local Similarity: 26.75% Mismatches: 934
 Query Match: 18,008 Indels: 1490
 DB: 18 Gaps: 147
 US-10-087-013-1 (1-10628) x AAM22475 (1-3060)
 QY 25 ATTATATAAATGGGATGCGAGCATCATCTATAGAGGAGATGCTAAAGCCCTATTATA 84
 |||
 Db 4 LeuAlalysketcllyProlysgluAlaAlaGlyGlyAspAspIleGluasp----- 20
 QY 85 AAAGAAAGTCACAAAGTCAGCAAGAAATGTTTGAACGTTATGCCAAATATATAGACAT 144
 |||
 Db 21 -----GluSerAlalysHISmetPheAspArgIleGlyLysAspValTyasp 36
 QY 145 CCATCAAAA---TATGCAAAAGAACATGTCGATTCGTGAAAGGATTTGAGCAAGCA 201
 |||
 Db 37 LysVallysgluGluAlaLysGluArgGlyLysGlyLeuIndGlyArgLeuSerGluAla 56
 QY 202 GAATTCGTCGTCGTCCTTACGCCAGTAATATAGCATATATATATTCATATCCA 261
 |||
 Db 57 LysPheGluLysAsnGluSerAspProGlnThrProGluasp-----Pro 71
 QY 262 TGTAAATTTAGATCATAGAACATCTAATTTAGCGTATGATGATGATTTGAGACAT 321
 |||
 Db 72 CysAspLeuAspHISLysTythrHISAsnVal-----ThrThrAsnValIleAsn 88
 QY 322 CCTTGCCATGCTAGAGCAAAACCGATTTGATGAAAGATCAACATCTGAATG---GGA 378
 |||
 Db 89 ProCysAlaAspArgSerAspValArgPheSerAspIleTythrGlyGlyGlnCysThrHIS 108
 QY 379 AATAAATATACGTAATATATAAAGAAAAATGATGCTATACCTGTGGCCACCTAGAGA 438
 |||
 Db 109 AsnArgIleLysAspserGlnGlnGlyAspAsnLysGlyAlaCysAlaProTyArgHIS 128
 QY 439 CGACATATGTGTATATAAAACCTTGAAGCTCTAATGATATTAATACCCAAATATTTCAT 498
 |||
 Db 129 LeuHISValCysAspGlnAsnLeuGluGlnIleGluProIleLysIleThrAsnThrHIS 148
 QY 499 GATTATATGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 558
 |||
 Db 149 AsnLeuLeuValAspValCysMetAlaAlaLysPheGluGlnIleSerIleThrGlnAsp 168
 QY 559 CATCA---CATAAAGCACT-----TCAGACCTGTACTGCTGTGCA 600
 |||
 Db 169 TyrProLysTythrGlnAlaThrTythrGlyAspSerProSerGlnIleCysThrMetLeuAla 188
 QY 601 CGAAGTTTGCAGATATAGTATGATGATGATGATGATGATGATGATGATGATGATGATG 660
 |||
 Db 189 ArgSerPheAlaAspIleGlyAspIleValArgGlyArgAspLeuTythrGluGlyAsnPro 208
 QY 661 CATGAC-----AAAGTGAAGAAAGGGTCTCGAAGGTTTCAGAAATA 705
 |||
 Db 209 GlnGluLysLysGlnArgGlnGlnLeuGlnAsnAsnLeuTythrIlePheGlyLysIle 228
 QY 706 CATGAT-----GGAATGGAAGATGAGTAAATAATGATATCAATCTGATGATCT 756
 |||
 Db 229 TythrGluLysLeuAsnGlnAlaGlnAlaArgTythrGlyAsnAsp-----ProGlu 244
 QY 757 GGAATTTATTTATTAATTAAGAGAACGATGCTGGAATGTGTAATGAATAATTAATGAGAA 816

```

Db 245 -----PhePheLysLeuArgGluAspTrpThrIleAsnArgGluThrValTrpLys 262
OY 817 GGTATATCATGATGATCATATTAATATGATATTTATGCAATCAAGAAATGATACA 876
Db 263 AlaiethrCysAsnAla---TrpGlyAsnThrTrpHe---HisAlaThrCysAsnArg 280
OY 877 CCATTAATTTCAATCCTAATGCGCCCATTAACAGAAAGTTCCTACCAATTTAGAT 936
Db 281 GlyIuArgThrLysGlyTyrCysArgCysAsnAspArgGluValProThrTyrPheAsp 300
OY 937 TATGTCCTCATATTTATGCTGGTTCGACAAATGGGAGACAGATTTGGCCAAAAGA 996
Db 301 TyrValProGlnTyrLeuArgTrpPheGlnIleTrpAlaGluAspPheCysArgGlyLys 320
OY 997 AATATTAATTTGAAAAAGTCAAGACTCCGTCTGT-----AATGCAAAAGA 1044
Db 321 AsnLysLysIleLysAspValLysArgAsnCysArgGlyLysAspLysGluAspLysAsp 340
OY 1045 CGCTTAATTTGCTACTCATATGACATGATTTGACAACTATTTGAAAAAGTATTT 1104
Db 341 Arg---TyrCysSerArgAsnGlyTyrAspCysGlyLysThrLysArgAlaIleGlyLys 359
OY 1105 TTGCATTTGCTATAGTACTGACTGTTCGACTAAATGCAAGTCTTTGAACTTTGG 1164
Db 360 LeuArgTyrGlyLysGlyLysIleSerCysLeuTyrAlaCysAsnProTyrValAspTrp 379
OY 1165 TTAGGAATCAACAGACATTTTAAACAAAGAAATATGAAAGAAATGAAATGCA 1224
Db 380 IleAsnAsnGlnLysGlnGlnPheAspLysGlnLysLysLysTyrAspGlnIleLys 399
OY 1225 TCATAT-----TTATCGAACGATTAACAAATTTGTCAT 1257
Db 400 LysTyrGluAsnGlyLysArgGlySerArgGlnLysArgAspAlaGlyThrThr 419
OY 1258 AATATTAATAGTGA---TTATTAACAATTTTATGAAGAAATGAAAGCAATAT 1314
Db 420 ThrThrAsnTyrAspGlyTyrGlyLysLysPheTyrAspLysLysSerGlnTyr 439
OY 1315 GCACATTAATGACCTTTTAAATTAATTAATGAAGAAAGTATTTGT----- 1362
Db 440 ArgThrValAspLysPheLysGlnLysLeuSerAsnGlnIleGlySerThrLysValLys 459
OY 1363 -----AAAGAGGATTTACAGAGAAAGAAAGATTAATCTTACTTAACAGTCT 1410
Db 460 AspGlnGlnGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 479
OY 1411 -----GATGCAAAAGGATTTTATCTGTCAGAAATATGCAAGTGTCT 1455
Db 480 SerGlyThrAsnValGluSerGlnGlyThrPheTyrArgSerLysTyrCysGlnProCys 499
OY 1456 CCCGACTGGGGGTCAATGATGATGATTAATAATACACACAC---AATCAGATTAATGAT 1512
Db 500 ProTyrCysGlyValLys-----LysValAsnAsnGlyLysSerSerAsnGln 515
OY 1513 CGTGAACGTGTAATTAAT-----GAAGACTATTAACCTTCACAGGGGT 1554
Db 516 TrpGlnGluLysAsnAsnGlnLysCysLysSerGlyLysLeuTyrGlnProAsp 535
OY 1555 GTGAAGCCCTACTAATATCACTGTCTTATAGTGTGAATGAACAGGTGATTTACACA 1614
Db 536 LysGlnGlyThrIleThrIleLeuLysSerGlyLysGlyHisAspAspIleGluIle 555
OY 1615 AATATGAAGAAATTTTGTGAAC-----AGCTCA 1641
Db 556 LysLeuAsnLysPheCysAspGlnLysAsnGlyAspThrIleAsnSerGlyLysSerGly 575
OY 1642 ACTAATTAACAAGATAAAAAAT-----CAAAATGGGAATGC 1680
Db 576 ThrGlyLysSerGlyGlyLysAsnSerLysArgGlnGluLeuTyrGlnIleTrpLysCys 595
OY 1681 TATTAAGGATGAAATATTAATAGATGTAACCTGCAACAAATATCAAGAAATCAATAT 1740

```

```

Db 596 ---TyrLysGlyLysAspValLysValGlyHisAspGlnAspArgGluAspTyr 614
OY 1741 GATATCTCTAG-----ATAATA----- 1758
Db 615 GluAsnValLysAsnAlaGlyGlyLeuCysIleLeuLysAsnGlnLysLysAsnGly 634
OY 1759 -----TCATTTGATTAATTT 1773
Db 635 GlnGlyGlyAsnThrSerGlyLysGluProAspGlnIleGlnLysThrPheAsnProPhe 654
OY 1774 TTTGAATTTAGGGTTCATATTTATTAAGGATTAATTAATGCAATGCAAACTTAA 1833
Db 655 PheTyrTyrTrpAlaIleHisMetLeuLysAspSerIleHisTrpLysLysLeuGln 674
OY 1834 ACTGTATA---AATATCAACACGACTGT---ATGTAATGTAAGAAATGTC 1887
Db 675 ArgCysLeuGlnAsnGlnLysAsnArgIleLysCysGlyAsnAsnLysCysAsnAsnAsp 694
OY 1888 TTATGTTTGACAGAGGTTTAAACAAAAGAAAGAAATGCAATGTAAGAAAGTGC 1947
Db 695 GluCysPheLysArgTrpIleThrGlnLysLysAspGlnTrpGlyLysIleValGlnHis 714
OY 1948 TTCACAAAAAAGAAATATACAG----- 1971
Db 715 Phe---LysThrGlnAsnIleLysGlyArgGlySerAspThrAlaGluLeuIle 733
OY 1972 -----CAATGCTATTAATTAATTAATTAATCTTTTGAAGTATTT 2016
Db 734 ProPheAspHisAspTyrValLeuGlnTyrAsnLeuGlnGluPheLeuLysGlyAsp 753
OY 2017 -----TTTAACTTAATGATTAACCTTGAACGAAAGTAAAGTGC 2058
Db 754 SerGluAspAlaSerGlnGlyLysSerGlnLysSerLeuAsnAlaGlnGluAla----- 771
OY 2059 AAAGACTTAAGAAATATTAAGAAAGAAAGAAATGAGTTTCCATTTGGAATATAT 2118
Db 772 -----GluGlnLeuLysHisLeuArgGlnIleIleGlnSerGlnAspAsn 787
OY 2119 AGGAGC-----TATTAGCAATGCAATTAATGCAATCTTGT 2154
Db 788 GlnGlnAlaSerValGlyGlyValThrGlnGlnLysAsnIleMetAspLysLeuLeu 807
OY 2155 GATCACTTAAAGAAACGCCACATATGT----- 2184
Db 808 AsnTyrGluLysAspGlnAlaAspLeuLysLeuLysIleHisGlnAspGlnGlnGln 827
OY 2185 ---AAGCAATTAATCAACAGCAATGCAATCCATTAATGCAACAAACCCG 2241
Db 828 LysGlnLysGlyAspGlyAsnGlnCysIleGlnGlnGlnLysGlnAsnPheArgTyrAsnPro 847
OY 2242 TGTGTTAAACCTCGTGGAGCAGCA---CCGCTAAATAATTAAGAAATGACAA 2298
Db 848 CysSerGlyGlnSerGlyAsnLysArgTyrProValLeuAlaAsnLysValAlaTyrGln 867
OY 2299 TACTTTAAAGAGTGCATACAGGAAGCAAGAAATCGTGTCTTCAATTAATGAAGA 2358
Db 868 MethHisLysAlaLysThrGlnLeuAlaSerArgAlaGlyLysSerAlaLeuArgGly 887
OY 2359 AAGCAGCAAGAGTATATTAACGTGGGGGTAGAGAAAGACTTCAAGCAATTA 2418
Db 888 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlyGlnIle 907
OY 2419 TGTGAATTAATGATTAACAACTTAATCGTAATCTTGTGTTCAATGACCATGTGAT 2478
Db 908 CysLysIleAsnGlnLysAsnTyrSerAsnAspSerArgGlyAsnSerGlyLysProCysThr 927
OY 2479 GCGAAAGCAGACAGTGTGATATCAACAAGATTTGCTGTAGCAATGGAAGTGC 2538
Db 928 GlyLys---AspGlyAspHisGlyGlyValArgMetArgIleGlyThrGlnTrpSerAsn 946
OY 2539 GATCCGGAACACATGGTAAAGATCAACGAAGATGTTATATGCTCTTGAAGAGCAGAT 2598
Db 947 IleGlnGlyLysLysGlnThrSerTyrLysAsnValPheLeuProArgArgGlnHis 966

```

QY 2599 ATATGTACATCCAAATTTGGAACATTTACAAACGATGATCCACCACTTAATGTAATAT 2658
:::|||||
Db 967 MetCysThrSerAsnLeuGluAsnLeuAspVal-----GlySerVal 980
QY 2659 GTT---GATGATTAGTAATTAATCTTTTGGGGGATGTTCTCTATCAGCAAAATAT 2715
:::|||||
Db 981 ThrLysAsnAspLysAlaSerHisSerLeuLeuGlnValGlnLeuAlaAlaLysThr 1000
QY 2716 GAAGCAACAACAGATATACGATATGATATTAAGAAAGATTAACCTTAAGGGCCCCAAAGAA 2775
:::|||||
Db 1001 AspaLalaIaGluIleIleLysArgTyrLysAspGlnAsnIle-----Gln 1016
QY 2776 GTAATCTGACCA-----AAACACAGACAACATCTGTCTGTGAGCTATACGTTACAGT 2826
:::|||||
Db 1017 LeuThrAspProIleGlnGlnLysAspGlnLysAlaIleMetCysArgAlaValAlaGlySer 1036
QY 2827 TTTTGACATATAGTGATATTAATTTGCGAGGAAGATCTCGGGAAGAAACGGTGACATG 2886
|||||
Db 1037 PheAlaAspLeuGlyAspLleIleArgGlyArgAspMetIrrAspGluAspLysSerSer 1056
QY 2887 GTAAGAGCTGCAAGACATTTGGAACCTTTTGGTATATACATAGTCACTCAAGAGC 2946
:::|||||
Db 1057 ThrAspMetCyluThrArgLeuIleThrValPheLysAsnIleLysGluLysHisAspGly 1076
QY 2947 --AAAGGAATGATTAATTAATGATGAT-----GCCCCCAATATTTAAATTTG 2994
:::|||||
Db 1077 IleLysAspAsnProLysTyrThrGlyAspGlySerLysProAlaTyrLysLysLeu 1096
QY 2995 AGGCAAAATTTGCGGAGCACTATAGCAAGCAAGTATGGAAGCAAGCATGAATGATATA 3054
:::|||||
Db 1097 ArgAlaAspIrrPrrGluAlaAsnArgHisGlnValIrrPrrArgAlaMetLysCys----- 1114
QY 3055 AAATATTTGAAGTAAATCGGGACACCAATCAACAAAGTATGTCGGATATAGT 3114
:::|||||
Db 1115 -----AlaThrLysGlyIleIleCys----- 1121
QY 3115 GATCATACACCATTTGATGATATATCCCAAAATTAAGATGATGACCGAATGGCA 3174
:::|||||
Db 1122 ProGlyMetProValAspAspTyrIleProGlnArgLeuArgTrrpMetCthrGlnTrrpAla 1141
QY 3175 GAATGAGTACGCAAGCTGCAAGCAAAAGAGATGATGATGATGGAAGCAAGCTGTAAGAG 3234
|||||
Db 1142 GluIrrPrrCysLysAlaGlnSerGlnLysIrrPrrAspLysLysIleCysAlaAsp 1161
QY 3235 TGTAAAGATTAAGATTAATGCTCAAGGCTGTACGAAAGAGAGTGTACAGGTTGTACAGAG 3294
:::|||||
Db 1162 CysMetSerLysGlyAspGlyLys---CysThr---GlnGlyAspValAspCysGlyLys 1179
QY 3295 TGCACAGACCTTGTATGATATATGATATGATATGATGATGATGAAAGCAACATGGAAT 3354
:::|||||
Db 1180 CysLysLalaIaCysAspLysTyrLysGlnGlnIleGluLysTrrpAsnGlnGlnTrrpArg 1199
QY 3355 ATATATACATTAATTAATCAAGATTTACAGCAACCAACAAATGCTGTAGTAATAGT 3414
|||||
Db 1200 LysLleSerAspLysTyrAsnLeuLeuTrrpLysGlnAlaLysTrrpThrSerThrAspPro 1219
QY 3415 GGTATTTGAAGCTTCAGTACTGCGCAAAATCATATAGAC-----AGGAATGTTAT 3465
:::|||||
Db 1220 Gly-----ArgThrValLeuGlyAspAspAspProAspLysGlnGlnMetVal 1235
QY 3466 GAATTTTGTGCGAATTAATCAACAAAGTGGGCAAA-----AGTAATTA 3513
:::|||||
Db 1236 AspPheLeuThrProIleHisLysAlaSerIleAlaLalaIrrValLeuValLysArgAla 1255
QY 3514 AGTGGTACTAGTGAATGAAATGCTGATGATGATGATGATGATGATGATGATGATGATGAT 3573
:::|||||
Db 1256 AlaGlySerProThrGluIleAlaAlaAlaAlaAlaProIleThrProLysSerThrAlaAla 1275
QY 3574 GCATATCTCATGATACAGCAAAATTTGATGATGATGATGATGATGATGATGATGATGATGAT 3633
:::|||||
Db 1276 GlyLysLleHisGlnGlnIleGlyTyrGlyCysGlnGlnGlnGlnGlnGlnGlnGlnGln 1295

QY 3634 GAAAA-----AGTATGCTAAGGATTAACGAAAAATATGCTTTAGA 3675
:::|||||
Db 1296 LysLleHisGlyAlaThrSerThrThrThrLysGlnLysGluLysGluTrrpPheLys 1315
QY 3676 GATTAACACAGACCATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3735
:::|||||
Db 1316 GlnProProGluLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1334
QY 3736 GTACAGTAAACGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGGAAAAAGG 3795
:::|||||
Db 1335 -----GluProLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1351
QY 3796 GATATCTTAAGAAAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 3855
:::|||||
Db 1352 LysIleLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1371
QY 3856 AGTAATGATATCCGATTTGCAATG---GAAATATTAATTTAGTGAAGACCTCGT 3912
:::|||||
Db 1372 -----TyrProAspIrrAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1387
QY 3913 GTGTATGCCCCCTAGAACACAAAGTTATGCGTACATTTCTGGCAATGATTAATGA 3972
|||||
Db 1388 AlaCysMetProProArgArgGlnLysLysCysLeuTyrIleAlaHisGlnSerGln 1407
QY 3973 ATAAAAAATTAACATCACAGTTAATTTAAAGAAAGCTTTCATCAATCTGCAGCAGCA 4032
:::|||||
Db 1408 ThrGlnAsnIleLysThrAspAspAsnLeuLysAspAlaPheIleLysTrrpAlaAla 1427
QY 4033 GAAACATTTCTTCATGCTTTATTTAAAGTAAAG---GATGCTGAAGAAATGAATCTC 4089
|||||
Db 1428 GluThrPheLeuSerIrrPrrGlnTyrLysSerLysAsnAspSerGluAlaLysIleLeu 1447
QY 4090 GATAAAGATTAAGAAAGCAAAATTCCTCCGCAATTTTGAATGATGATGATGATGATGATGAT 4149
:::|||||
Db 1448 AsparG-----GlyLeuIleProSerGlnPheLeuArgSerMetMetTyrThr 1463
QY 4150 TTTGACATTAATGAGATTTTATTTTGAACAGATTAATCAAAAGCTGATGATGAGCA 4209
:::|||||
Db 1464 PheGlyAspTrrArgAspIleCysLeuAsnThrAspIleSerLysGlnAsnAspVal 1483
QY 4210 AGTAACCTTAAGCAAGCAATTAATGATTCCTTTGCAAAATGCTGACCAAAATCTCTAAT 4269
:::|||||
Db 1484 AlaLysLysAspLysLysLysLysLysPhePheSerLysAspLysLysSerProSer 1503
QY 4270 GGAAGAAACAGCCAGCAAGTGTGAGAGCAACATAGCATGATATGGAAGTATGCTA 4329
:::|||||
Db 1504 GlyLeuSerArgGlnGlnIrrPrrLysThrAsnGlyProIleLysLysLysLysLysLys 1523
QY 4330 TGTGCTACTTAATAATTTGGGCAAAAAAGATGAT-----TTTACCGAAAACCTAC 4380
:::|||||
Db 1524 CysAlaLeuThrLysTyrValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr 1543
QY 4381 GGTTAACAACAAGCTCAATTTAGTACAAA---AGCAACCTTTGAGGAATTTGGCAAA 4437
:::|||||
Db 1544 SerTrrpLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnGlnPheAlaAla 1563
QY 4438 CGACCCAGCTTTTACGATGCTTACCGAATGCTTACGACATGCTTATACAGCAAA 4497
:::|||||
Db 1564 LysProGlnPheLeuArgTrrpMetIleGlnTrrpGlyGlnLysLysLysLysLysLysLys 1583
QY 4498 AAATATTTGAAGATGTCAGCAAAATGTAAGTCA---AATGACCAATTTGAAGTGTAT 4554
:::|||||
Db 1584 LysLysGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1603
QY 4555 ACA-----GAATGATTAAGAAATGCGAGACTACTTAAATATATAGAAA---AAA 4602
:::|||||
Db 1604 AspaLalaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnLysValGluAsnLys 1623
QY 4603 AAAAAAGAGGATTTCCACCAAGATTAATTAATCAAGATGATGATGATGATGATGATGATGAT 4662
:::|||||
Db 1624 LysLysGlnPheSerGlnGlnIrrPrrAsnAsnPheValLeuLysLysLysLysLysLysLys 1643
QY 4663 GATAGACACACATGCTGTAATGTTTACAGACTATTAATGGAAGCAATGCAAGATTAAC 4722

```
Db 1644 AspProGluTyrLysGlyTyrGluTyrLysAsp----- 1654
QY 4723 TTGAACAGCAAAATTACTGTAGTTGGTGTATAGCCTGGAAAGTCCCTCTGTGTACAA 4782
Db 1655 -----GlyValGlnPro-----IleGln 1660
QY 4783 AGAAATATACAAATTGTAGAAAAACAGGCTTACTATGTATGCCGCAAAACATTTGGGTGC 4842
Db 1661 GlysngluTyrLeuGlnLys-----CysaspasnLys---CysSerCys 1676
QY 4843 ACAAATTTATTTGAATGACGACAAATATCTAATTTGAGTAAAGATAGTGCAA 4902
Db 1676 ----- 1676
QY 4903 GGATTAGTAAAGAGCAACACAGAGTGTATTAAGTGCAAAACAAGGTCCTAATAC 4962
Db 1676 ----- 1676
4963 TACAATACCTTGAAAGAAATTGACTGAAGATGTGCTTTCTCTCGTCGACTACGTATA 5022
1676 ----- 1676
QY 5023 TGTTCATGATGTGATGGCAATTATACAGATCCAGAAATGAAGTGAATGGGTG 5082
Db 1677 -----MetaspGlyasnValLeuSerValSerProLysGluLys----- 1689
QY 5083 CGAAAGATTTGATGAAAGTGGCGGCAACGAAAGGTACAAATTTGGGTCAATCTACAAA 5142
Db 1690 -----ProPheGlyLysTyr----- 1694
QY 5143 GAAAAAGAAAAAGAAAAATTAAGTGGATGCGGCAAAATTTCTTATGAGGTC 5202
Db 1695 -----AlaHisLysTyrPro----- 1699
QY 5203 CCOCCTTGATGCTATGAATATAGTTTATGATTGAAGATATTAATCTAGTAT 5262
Db 1699 ----- 1699
QY 5263 GATAAATTTGSAAGATGAAAAACAAGACCCGAGAAATTTTGAGAAAAATTTTAACAAA 5322
Db 1699 ----- 1699
QY 5323 AATGGAACATCACTGTGCAAGAGAGTATGATCTACAGAAATCCGGTACTGCG 5382
Db 1699 ----- 1699
QY 5383 CGAAATTTTCTGAGAGAAATAAGAAATGTGTGTGAGACGCAATGATGCGGTAC 5442
1699 ----- 1699
QY 5443 AAACGTGTAGGAGATGATGAAATAGTGAATAGTGCAGAAAGATGATGAATCTAAAA 5502
Db 1699 ----- 1699
QY 5503 AAATGTGTTCTGTACCTTCAGATGATGATTTCTATGAGGAAAAATCGCATGAAGT 5562
Db 1699 ----- 1699
QY 5563 ACTGGTATCACTTCTTCGATGATGTTTGGCCGATGGGGAAGATTTTGCAMACATAAA 5622
Db 1699 ----- 1699
QY 5623 GAAAGGAATGGAGAAATTTGTTAGGCGGTATGATTAATCTTGTTGGTATATGAA 5682
Db 1699 ----- 1699
QY 5683 GATTAAGAAAGAAATGTACAGATCGGTACACAAATATTAATAATTTATAGTAGTGG 5742
Db 1700 -----GluLysCys----- 1702
QY 5743 AAACCAAGATATGAAAAACAATCAAAAAATATGTGTGGAATTAAGACAAAAATATATCC 5802

Db 1702 ----- 1702
QY 5803 GACATCTCTGTGGCAAAAGATGACAGAGACGCTCGGCAATATTTAGCAACAATTAATA 5862
Db 1702 ----- 1702
QY 5863 AAAATTTGAAAAATTAAGATGAGATTTGTAATATAGTATGAAAGATGTGCCACA 5922
Db 1703 -----AspCys----- 1704
QY 5923 CACGATTAACGTATGATAGTCAAAATATGCCGATCATTTAGAGATGAACAAAA 5982
Db 1704 ----- 1704
QY 5983 GAAATTGAAGAAAGTATTAATTTGTCAAGTCCGACGAGGTCCACAGCTGTACAGAGGA 6042
Db 1705 -----TyrGlnGlyLys-----HisValProSerIleProPro----- 1715
QY 6043 ACACGCTCACACAGGCTATCACTGATATCAAAAAGCAGCGCATCGAAAAAGACGAAA 6102
Db 1715 ----- 1715
QY 6103 ACAGCGCCGCTACAAAACAGCCGAAAAAGTGAAAAATCTAACACAGAAATGCGACA 6162
Db 1716 -----ProProProProValGlnProGln----- 1723
QY 6163 CAACACGACGCCGACGACGACGACACACACACGAAAAACGAACTCAACAGCAACA 6222
Db 1724 -----ProGluAlaProThr----- 1728
QY 6223 ACAGATCTGACGTGGGACACATGTGTAAGGCAATCTTTCGAAATTAACAGATAGCAG 6282
Db 1729 ValThrValAspValCysSerIleValLysThrLeuPheLys-----AspThrAsn 1745
QY 6283 GGTGAATAGAGCGTTGTATATCCAAAAACGTAATGACAAAT-----CTAAATGGCGT 6336
Db 1746 AsnPheSerAspAlaCysGlyLeuLys-----TyrGlyLysThrAlaProSerSerThrLys 1764
6337 TGTAT-----GTAAGTAAGTCTAAAGAAAT 6363
Db 1765 CysIleProSerAspThrLysSerClyAlaGlyAlaThrThrGlyLysSerGlySerAsp 1784
QY 6364 GAAATGCAATATGATGCTCTCTAGGAAAAAAATTTATGATTAATATACATAT 6423
Db 1785 SerGlySerIleCysIleProProArgArgArgLeuValGlyLysLeuGln--- 1803
QY 6424 TTAATATATGAAC-----GAAATACAGCTGAC 6453
Db 1804 -----GluThrPalatThrAlaLeuProGlnGlyGluAlaAlaProSerHisSerArgAla 1822
QY 6454 AATGATATTAAGAGCGCTTTTATTAATAGTGCACCAATAGAACTCAATTTTGTGTTA 6513
Db 1823 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuThrPasp 1842
QY 6514 AAATATATATGAA-----AATCTGCACAGAAAAATGAA----- 6549
Db 1843 ArgTyrLysGluGluLysLysProGlnGlyAspLysSerGlnAlaLeuSerGlnLeu 1862
QY 6550 -----TTGCAAAATGACACA 6564
Db 1863 ThrSerThrTyrSerAspAspGluGluAspProProAspLysLeuGlnAsnGlyLys 1882
QY 6565 ATTCCAGATCAATTTAAAGAAATATATGATTAATACATATGATTAATGAATGATTT 6624
Db 1883 IleProProAspPheLeuArgLeuMetPheTyrThrLeuGlyLysPyrArgAspIleLeu 1902
QY 6625 TTGGAACGTATATTTCTATGATTAATAAAATATATATACCTATACCAATAGCTGAACACC 6684
Db 1903 ValHisGlyLysAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1922
QY 6685 ATTCTCAATGAATAATATTAAGAAAAACAGATAAAAAAACAGATAGATTAAGCTATAA 6744
Db 1923 ValLeuGluLysSerGlyAsnLysGluAspMetClnLysIleGlnGlyLysIleGln 1942
```

QY 6745 ATA----- 6747
Db 1943 IleuProLysaenGIyGlyThrProLeuValProLysSerSerIaGIInThrProAsp 1962
QY 6748 ---TTTGGGAGAAAAATTTAAATTTATTTGGAGAGATGATATATGATTACTTAT 6804
Db 1963 LysTrpTrpAsnGluHisIaIaGluSerIleTrpLysGluMetIleCysAlaLeuThrTyr 1982
QY 6805 -----CATCTCAGACGAGAAAAACGAAAAAGAAAAATATAGA 6840
Db 1983 ThrGluLysaSnProAspThrSerAlaArgGlyAspGluAsnLysIleGluLysAspAsp 2002
QY 6841 GATTAATTAC----- 6849
Db 2003 GluValLysGluLysPhePheGlySerThrAlaAspLysIleGlyThrAlaSerThrPro 2022
QY 6850 -----CAGTACAT---GACATGACCAACTG----- 6873
Db 2023 ThrGlyThrTyrLysThrGluTyrAspTrpGlyLysValLysLeuGluAspThrSerGly 2042
QY 6874 -----ACGCTTCCTTGAGAGGTTTGTAAAAAGGCC 6906
Db 2043 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheValLeuArgPro 2062
QY 6907 CAATTTTGAAGATGTTACAGAAATGGCAGAGAAATTTGTATATAGAGAGAGAAAG 6966
Db 2063 ProLysPheArgTyrLeuGluGluGluTrpGlyLysAsnPheCysLysLysArgLysHisLys 2082
QY 6967 TTGTTAAATTTGAGGGGGCTGTAAAGAAATATAG- 7002
Db 2083 LeuAlaGluIleLysHisGluCysLysValGluGluAsnGluGlyLysArgArgGly 2102
QY 7003 -----TGTAAT-----GGTACTAAT 7017
Db 2103 GlyLLeuThrArgGluTyrSerGlyAspGlyGluAlaCysaSnGluMetLeuProLysaSn 2122
QY 7018 GAGCGTAAGACACAAAGAA-----TGTCAGAGGGCTGTACATATCA 7062
Db 2123 AspGlyThrValProAspLeuGluLysProSerCysAlaLysProCysSerSerTyrArg 2142
QY 7063 AATTTTAAAGAGTGAAGAACTGATATGAAGACAAAGAGAAAGAAATTCMAAAAGAT 7122
Db 2143 LysTrpIleGluSerLysGlyLysGluPheGluLysGluGluLysAlaTyrGluGluGlu 2162
QY 7123 AAAGAT-----GGCAAAAAGTATAGGATTATCTTCTACTGAAGAGACATA 7170
Db 2163 LysAspLysCysValaSnGlySerAsnLysHisAsp----- 2174
QY 7171 GAGAGGCAACATGTGCTCATGAATATTAAACATGAATTAAGAAATTAATGTGCAAT 7230
Db 2175 -----Asn 2175
QY 7231 AAGGATGTCTTGATGCAAAAACCTTCTCACAACTACCAAAAACACAAACATCA 7290
Db 2176 GlyPheCysGluThrLeu-----Thr 2182
QY 7291 CAATCATCCGATCGTAATGATATCCAGAAATCCGTGATTAATCTCTGAAGAAATTAAC 7350
Db 2183 ThrSerLysAlaLysAspPheLeuLysThrLeuGly----- 2195
QY 7351 AAGGTGAGTCTGTAACCTTCAAAAAAGGATCTATGAT 7392
Db 2196 -----ProCysLysProAsnAsnValGluGlyLysThrIlePheAspAspLysThr 2213
QY 7393 -----CATCAAAAAAATTTACTGAACT-----AAATACTATGATTTGTGAGAG 7440
Db 2214 PheLysHisThrLysAsp---CysAspProCysLeuLysPheSerValaSnCys----- 2230
QY 7441 AAACGACATATTTATCTAAGAGAGCAAGAAATATATGATATTACTCTGAAGAA 7500
Db 2230 ----- 2230

QY 7501 AAATTTATACCTATTAGTGTACAAAGAGAAAGAAAGTAAATAGTTGACTAATAT 7560
Db 2231 -----LysLysaSpGluCysaSpaSnSer----- 2238
QY 7561 AATCCTTCGATCTCTAAGAAACCTTATGCACCTGATTAATATATATAGACAAAGAACCT 7620
Db 2239 -----LysGlyThrAsp 2242
QY 7621 TGTCAAAATAGAGAGAAATCGTTTAAGGTAGATTATGAAATGAAATGTTCAAAAAT 7680
Db 2243 CysArgAsn-----LysaSn 2247
QY 7681 TCAAAAGTCTATCAGAGAGAAAAAGAGTATGTGTACTCCAAAGAGAACATATGTGC 7740
Db 2248 Ser----- 2248
QY 7741 TTAAGAAATTTAGATGAATTTAAATTTGAAAGACTTAAGATAGTAAATTTCTCTAAA 7800
Db 2249 -----IleAspAlaThrAspIleGluAsnGlyValaSpSerThrValLeuGluMet 2265
QY 7801 ATGGTTCGCACTGCAGCAAAATGAAGAAATAGACATATATAAAAACTTCAACTCA 7857
Db 2266 ArgValSerAlaAspSerLysSer-----GlyPheAsnGlyAsp 2278
QY 7858 -----GAGAACGGGTGCGCAATGAATCAATATGTGATCTATGAATATAGTTTGGCT 7911
Db 2279 GlyLeuGluAsnAlaCys----- 2284
QY 7912 GATCTGGGTGACATAGTTAGAGCAACAGATATGTTCAGAAATTTGGTGTACTTACCTCC 7971
Db 2285 -----ArgGlyAlaGly----- 2288
QY 7972 GTAGAAATTAATTAATTAAGTGTGTTTGAATACATATATGAAATATGAGAAATTAATAAT 8031
Db 2289 -----IlePheGluGlyLe----- 2293
QY 8032 AAAGGTAAATTAATTAATTAAGATGTACAAACCTTCTGCTGTGGTGGATCTAAT 8091
Db 2293 ----- 2293
QY 8092 AGAAAGATATTTGCAAGCA-----ATGACGTGCAAGACCGAGAA 8133
Db 2294 ArgLysaSpGluTrpLysCysArgAsnValCysGlyTyrGluValCysLys---ProGlu 2312
QY 8134 GATGCAAAACCTTTTACAAAGAGAGAAAGATGATGATTTGACGATTAATTAATCAAA 8193
Db 2313 AsnValaSnGlyGluAlaLysGlyLys----- 2321
QY 8194 GATAGTGTGACATTAAGAGCATCCACTGTGATGATTAATATACCTCAACGGTTTGA 8253
Db 2322 -----HisIleIleGluIleArgAla 2328
QY 8254 TGGATGACGTGAAGCTGTGAATATATTGTAAGSCACTGATGGAAGAAATGGAATAATTT 8313
Db 2329 LeuValLysArgTrpValGluTyrPhe-----PheGluAspTrpAsnLysIle 2344
QY 8314 AAAAATCATGTGATCACTGTAAACATCTGACAGATGCAAGATGATGATGAAGAAAT 8373
Db 2345 LysHis-----LysIleSerHisArgIleLysaSn-----Gly 2355
QY 8374 AAGGTGACAGCTTAAGAGAGATGTCAAGATATATAAATTTGTTTAAATGAA 8433
Db 2356 GluIleSerProCys-----IleLysaSnCysValGluLysTrp--- 2368
QY 8434 TCTCTATTCGATATACAAATCAATTAATTAAGAAATGTATGATGACCAACCAATATATACA 8493
Db 2369 -----ValaSpGluLysArgLysGluTrpLysGluIleThrGluArg----- 2382
QY 8494 AAATCTCTACTTATGATCATGTTCAAAATTTGTACAAAAGTTGAAACCTTTAAAGT 8553
Db 2383 -----PheLysAsp 2385
QY 8554 GAATGTTCTGTGAGACCTTTTGAATATCTTCATGAAACAAAGTAAGTGTGAATTAAT 8613


```
Db 2386 Gln -----Tyr 2387
QY 8614 AAATTAATGAAGATGATGTTCTTCCAATATAGCAACATATGCTTCGAAAGAACCA 8673
Db 2388 Lys ---AsnAspAsnSerAspAspAsnValArgSerPheLeuGlnThrLeuIlePro 2406
QY 8674 AAA-----ACTTATAAGAAAGCT 8691
Db 2407 GlnIleThrAspAlaAsnAlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSer 2426
QY 8692 TGCAGTTGTACACCTCTCTCAAGATCCATGATATTTGCTCCACCATCAAAACAA 8751
Db 2427 CysGlySerAlaSerAla-----AsnGlnGlnAsnLys 2438
QY 8752 GATGAGTAAAGAAATTACAACCTTACCTCTGCTCGAAGAAATGATTATTAATAT 8811
Db 2439 AsnGly----- 2440
8812 CTGTATTAATTGGAAGCATTACCTGTTCTTAATATGTCAGATGATACAAAGGTATATG 8871
2440 ----- 2440
QY 8872 ATTCCTCAAGAGAACACATTTATGTACAAAGCATTACCTGATATATTTAGAAA 8931
Db 2440 ----- 2440
QY 8932 GGTGATTAAGAAATTTTAAAAAACTTCTTACTTCTGCTTTCAGTCAAGACAATTG 8991
Db 2440 ----- 2440
QY 8992 TTAGTCAAAATATTAATGCGAAGAAAGCTGCTGTTGAGCAATGAATATATCTTAT 9051
Db 2440 ----- 2440
QY 9052 GCAGATTATTCGATATATTAAGAACACATGATGATGACACATTCATTATCTGA AAA 9111
Db 2441 ---GlnTyrLysAspAlaIle-----AspCysMetLeuLysLysLeuLysAspLys 2456
QY 9112 ATTAAAAAAATATTTGAACATCAATTAAGAACACGAAAAATCGTAAACATGTTGGAA 9171
Db 2457 Ile----- 2457
QY 9172 AATAATAGACGTCAGATATGCGACGCTATGTATGTGATATTAATTTGCTACTTCAAAA 9231
Db 2457 ----- 2457
QY 9232 GTAACTATTAAGTAAGAGATGCTGTCAATTACCAAGATGAAGAACTAATCAGTTCTT 9291
2458 -----GlnGlnLys----- 2460
QY 9292 CGTTGTTAATTGAATGGGCAAGACATGTAAGAAAAAACATGTAAGTATCA 9351
Db 2461 -----GlnLysLysHisHisGlnThrSer 2468
QY 9352 TTAACCAACAAATGT-----CCGCTTCAACGACGATATTTTGAAGCG 9396
2469 ---AspThrGlnCysSerAspThrProGlnProGlnThrLeuGlnLysP----- 2483
QY 9397 TCAGAAATTATTAAGACACCTGATGTCAGAAATGATTAAGAAATATATTAAGCTTGAAT 9456
2484 ---GlnThrLeu-----AspAspAspIleGlnThrGlnGlnLysLys 2497
QY 9457 ATATGATTAATAATACAAATGGAATGAAATTAATTAATTAATTAATTAATTAATTAATTA 9516
Db 2498 AsnMetMetProLysIleCysGlnLysVal---LeuLysThrAlaGlnGlnLysPglu 2516
QY 9517 TCTTCAGATTAATATAGACAATTAACCATCTGACAGAAAATGTTCACTCATATTAATAATCA 9576
2517 ---GlnGlyCysValProAlaGlnLysSerGlnGlnLysPglu-----Ala 2530
QY 9577 AAAGATTCTCAATGCGCTTGGAGTTAAATGATTAATTAATTAATTAATTAATTAATTAATTA 9636
|||||
```

```
Db 2531 ThrAspSer----- 2533
QY 9637 AATAATGAATAATGAATTCAAAGAGTACTAAAAAATTAATATCTGTTATATTT 9696
Db 2534 GlyLysGlnThrProGlnGlnThrProValLeuLys----- 2545
QY 9697 GTTGAAGATGAACACACAAAAATCATGTACTAGATGAAATTAAGAAAGACAGCA 9756
2546 -----ProGlnGlnGlnAla 2550
QY 9757 ACAGTTCGCTTAAGACACTATTTCTTACACCCCATGATGATTTCTTATACAGA 9816
Db 2551 ValProGlnProPro-----ProProProGlnGlnLysAlaProAla 2565
QY 9817 CCTTTATTCACACATGACATGACACAATATGATCTTAAAAAT----- 9861
Db 2566 Pro-----IleProGlnProGlnProProThrProThrGlnLeu 2579
QY 9862 -----GATTAATTAAGAAAGTACTATCTGCTGTTATGTATGCGGCTTAGCT 9909
2580 LeuAspAsnProHisValIleuThrAlaLeuValThrSerThrLeuAlaThrSerValGly 2599
QY 9910 TTTG-----ATACGCTTCATTCATGAAAGAAAAATTCAAATCGCTGTG--- 9954
Db 2600 IleGlyPheAlaThrPheThrThrPheThrLysLysLysThrLysSerSerValGly 2619
QY 9955 GACTGTTCCGATATCTGAAATATCCCGAAGAGATATGAAATCGCTTGGAAATCG 10014
Db 2620 AsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThrLysLeuSer 2639
QY 10015 AAAAATAGTATACATACATATATGAAGTGCATTAAGGCAAAAAACATATATATATAT 10074
2640 ProAsnArgTyrIleProThrThrSerGlnLysTyrAlaGlnLysArgTyrIleGlnLeu 2659
QY 10075 GAAGAGATACI-----AGTGAGATGAAGATTAATATATGTTGGACTTATCTTCC 10125
2660 GlnGlyAspSerGlnLysAspSerGlnLysThrAspHisTyr----- 2673
QY 10126 TCTGATTAATTAATTCGCAAGATGATGAAATGATTAATTAATTAATTAATTAATTAATTA 10185
2674 SerAspIleThrSerSerGlnSerGlnLysGlnLysGlnLysGlnLysGlnLysGlnLysGln 2693
QY 10186 CCAGTACTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 10245
2694 ProGlnSerProLysTyrLysThrLeuIleGlnValIleGlnLysProSerGlnLysAsn 2713
QY 10246 ATACCAAGTAT----- 10257
2714 ThrThrAlaSerGlnLysAsnThrThrAlaSerGlnLysAsnThrThrAlaSerGlnLys 2733
QY 10258 GATACACCAAGT-----AATGATACACACGATGACGATGATGATTT 10296
2734 AsnThrProSerAspThrGlnAsnAspIleGlnLysAsnAspGlyIleProSerSerLysIle 2753
QY 10297 ATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10347
2754 ThrAspAsnGlnLysPheGlnLysAspGlnLysAspGlnLysAspGlnLysAspGlnLys 2773
QY 10348 CCAATACAGAACCAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10407
2774 ProAsnThrGlnPro-----AsnMetLeuGlnLysAsnAspAsnAsnThrHisPro 2791
QY 10408 AATACCTTAATTTCTGATTAATCTGAAAGAAAAACCTTTATTAATTAATTAATTAATTA 10467
2792 ThrThrSerHis---HisAsnValGlnGlnLysProPheHisMetSerIleHisAspArg 2810
QY 10468 GATTATTAATCTGGAAGAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 10497
2811 AsnLeuPheSerGlnGlnLysGlnLysAsnTyrAspMetPheAsnSerGlnLysAsnProIle 2830
QY 10498 AATATTAATTAATGATTAAT-----ACTAATAATGATATTAATTAATTAATTAATTA 10539
2831 AsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsnHisSerProTyr 2850
```


QY 10540 AATGCTAGAAATGATCTTATAGAGGATATGATTTAATGATTCACCTAGTGGTCTA 10599
DB 2851 AsnAspIysAsnAspLeuTyrSerGlyIleAspLeuIleAsnAspAlaLeu--SerGlyA 2870
QY 10600 AACCTATGATATATATGATGAGCAATGATG 10628
DB 2870 snHisIleAspIleTyrAspIleuMetLeu 2879
RESULT 3
AAV77905
ID AAV77905 standard; Protein: 3060 AA.
XX
AC AAV77905;
XX
DT 13-JUN-2000 (first entry)
XX
DE Plasmodium var-7 polypeptide.
XX
KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
XX
AX DBP; Stialic Acid Binding Protein; SABP; malaria; vaccine; Immunisation;
XX
OS Plasmodium sp.
XX
PN US5993827-A.
XX
PD 30-NOV-1999.
XX
PE 07-JUN-1995; 95US-0487826.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI SIm KL, Chitins C, Peterson DS, Su X, Wellem TE, Miller LH;
XX
DR WPI: 2000-194198/17.
XX
DR N-PSDB: AA298287.
XX
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
XX
PT falciparum erythrocyte binding proteins useful for vaccinating against
XX
PS malaria -
XX
PS Disclosure: Columns 109-124; 93pp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
XX Binding Protein (SABP), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABP are the
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
XX CC and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX CC Immunization with the polypeptide provides effective protection against
XX malaria. The present sequence represents the var-7 polypeptide.
XX
SQ Sequence 3060 AA:
Alignment Scores:
Pred. No.: 5.39e-238 Length: 3060
Score: 3501.00 Matches: 1057
Percent Similarity: 38.65% Conservative: 470
Best Local Similarity: 26.75% Mismatches: 934
Query Match: 18.00% Indels: 1450
DB: 21 Gaps: 147
US-10-087-013-1 (1-10628) x AAV77905 (1-3060)
QY 25 ATATATAAATGGGAGATGACATCATATAGAGGAGATGCTAAAGCCCTATTATA 84
DB 4 LeuAlaIysMetGlyProIysGluAlaIleAlaIleGlyIAspAspIleGluAsp----- 20

QY 85 AAGAAAGTCACAAAGTCAGCAAAATGTTTGGACGCTTATGCGCAAAATATPAGACAT 144
DB 21 -----GluSerAlaIalysHisMetPheAspArgIleGlyIAspValTyrAsp 36
QY 145 CCATCAAAA---TATGCAAAAAGAACATGTGATTCGTTGCAAAAGCGATTTGACAAAGCA 201
DB 37 LysValIylsGluGluAlaIalysGluArgGlyIylsGlyLeuGlnGlyArgLeuSerGluAla 56
QY 202 GAATTCGTCGTGTCCTCTTACGCGCAGTAAATGACATATATATATATATATATATATCA 261
DB 57 LysPheGluIylsAsnGluSerAspProGlnThrProGluAsp-----Pro 71
QY 262 TGTATATTAGATCATTAAGAAACATATCATTAATTTACGTATGATGATGCAATTTGACACAT 321
DB 72 CysAspLeuAspHisIylsTyrHisThrAsnVal-----ThrThrAsnValIleAsn 88
QY 322 CCTGGCCATGTCAGCAAAACCGCTTTGATGATGATGATGATGATGATGATGATGATGATG 378
DB 89 ProCysAlaAspArgSerAspValArgPheSerAspIulTyrGlyGlnCysThrHis 108
QY 379 AATTAATATAGTAAATTTATTAAGAAAAAATGATGCTTACGCTGCGCCACCTAGAGA 438
DB 109 AsnArgIleIylsAspSerGlnGlnGlyAspAsnIylsGlyIalCysAlaIleProTyrArg 128
QY 439 CGACATATGTGTGATTAATAAAGCTTGAAGCTCTTAATGATTAATATATATATATATCAT 498
DB 129 LeuHisValCysAspGlnAsnLeuGlnIleGluProIleIylsIleThrAsnThrHis 148
QY 499 GATTTATTTGGAAATGTACATGATACAGCAAAATACAGAGTGAATCATATTTATTAAT 558
DB 149 AsnLeuValAspValCysMetAlaIalysPheGlnGlnSerIleThrGlnAsp 168
QY 559 CATCA---CATAAAGGACT-----TCAGACGCTTACTGCTCTGCA 600
DB 169 TyrProIylsTyrGlnAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 188
QY 601 CGAAGTTTTCGACATATATGATATTTGTAAGCAATAGATATGTTTAAACCAATGTC 660
DB 189 ArgSerPheAlaIalysIleGlyAspIleValArgIylsArgAspLeuTyrLeuGlyAsnPro 208
QY 661 CATGAC-----AAGTAGAAACGGCTCGGAGAGGTTTTCAGAAATA 705
DB 209 GlnGluIleIylsGlnArgGlnGlnLeuGlnAsnLeuIylsThrIlePheGlyIylsIle 228
QY 706 CATGAT-----GGAATGAGATGATGAATTAATAATGATTCATTCCTGATGATCT 756
DB 229 TyrGluIylsLeuAsnGlyAlaIalArgTyrGlyAsnAsp-----ProGlu----- 244
QY 757 GGAATTTATTTATTAATTAAGAAACATGCTGCAATGTGATGATTAATTAATGCGAA 816
DB 245 -----PhePheIylsLeuArgGluAspTrpIlePheAlaAsnArgGluThrValTrpLys 262
QY 817 GCTAATACGATGATGATGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 876
DB 263 AlaIleThrCysAsnAla---TrpGlyAsnThrTyrPhe---HisAlaThrCysAsnArg 280
QY 877 CCATTTATTTCAATCTTAATGCGGCGCATTAACAGAAAGTTCCTTACCAATTTGAT 936
DB 281 GlyIleuArgThrIylsGlyTyrCysArgCysAsnAspAspGlnValProIleThrPheAsp 300
QY 937 TATGTCCTCAATATTTTACGTTGTTGACAGCAATGGGAGAGAGATTTTGGCAAAAAGA 996
DB 301 TyrValProGlnTyrLeuArgTyrPheGlnGluIulTrpAlaGluAspPheCysArgIyls 320
QY 997 AATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1044
DB 321 AsnIylsIylsIylsAspValIylsArgAsnIylsArgGlyIylsAspLysGluAspLysAsp 340
QY 1045 CGCTTATATGTCATTAATGACATGATGTACGCAACTATTTGAAAAAGGTAT 1104
DB 341 Arg---TyrCysSerArgAsnGlyTyrAspCysGluIylsThrIylsArgAlaIleGlyLys 359


```
QY 5023 TGTTCATGCATTTGGATGCAATTTATACAGATCCAGAGTTGAAGATGAATGGGTG 5082
Db 1677 -----MetaspGlyAsnValLeuSerValSerProLysGluLys----- 1689
QY 5083 CGAAAAAGATTGATGGAAGTGGCGCAACGAGAGGTCAATTTGGGCAATACACAAA 5142
Db 1690 -----ProphegLyLysTyr----- 1694
QY 5143 GAAAAAAGAAAAAGAAAAATATAAACGTGGATGGCGCAAAATTTCTTATAGAGTTC 5202
Db 1695 -----AlaHisLysTyrPro----- 1699
QY 5203 CCGCGTTGTAGTGTATGAATATAGTTTTTATGATTTAAGATATTAATTTAGTATT 5262
Db 1699 ----- 1699
QY 5263 GATAATTTGGAAGATGAAAAAACAAGACGAGGAAATTTGAGAAAAATATTAAACAA 5322
Db 1699 ----- 1699
QY 5323 AATGGAACATCAGTTGGCAAAAGAGATGATAGTACTACAGAAATCCCGGTAGTACTGCG 5382
Db 1699 ----- 1699
QY 5383 CGAAAAATTTTTCGGAACGAAAAATTAAGGAATGTGTGGAACCGCAATGATATCGGGTAC 5442
Db 1699 ----- 1699
QY 5443 AAACGTGTAGGAGATGAGAAATAGTGAATATGTCAGAAAGTGTAGAAAGATCTAAAA 5502
Db 1699 ----- 1699
QY 5503 AAATGTGTTCTGTACTTCTAGATGATGATTATTCCTATGGGAAAAAATCGCGATGAAGT 5562
Db 1699 ----- 1699
QY 5563 ACTGCGATCAGTTCTCGATGTTGCCGAAATGGGATGAAGATTTTTCGAAACATAAA 5622
Db 1699 ----- 1699
QY 5623 GAAAAAGAAATGGAATAATGTTAGGCGGTGAATGATTACTTGTGTATATGAA 5682
Db 1699 ----- 1699
QY 5683 GATTAAGAAGAAATGTACAGATGCGGTACACATATATAAAATTTATAGTGTG 5742
Db 1700 -----GluLysCys----- 1702
QY 5743 AAACCACTATGAAAAACAATAATGATGAGATAAAGACAAAAATATATTC 5802
Db 1702 ----- 1702
QY 5803 GAGCATCTGTGGCAAAAAGATCAGAGACGCTCGCGAATATTAGACAAACATTTAAA 5862
Db 1702 ----- 1702
QY 5863 AAAATTTGTGAATAAAGTGAAGATTTGTGAATATAGTGTAGAAAGTGTCCACA 5922
Db 1703 -----AspCys----- 1704
QY 5923 CAGCATTAACATGATGATTAATAGTCAAAATATGCCGATCATTTAGACATGAACAAA 5982
Db 1704 ----- 1704
QY 5983 GAAATTTGAAGAAAGTGAATTTGCAAGTGCACAGAGTCCACACGCTGTACGAAGGAA 6042
Db 1705 -----TyrGlnGlyLys-----HisValProSerLLeuProPro----- 1715
QY 6043 ACACCGTACACACGGGTATCACTGATATCAAAAGCGACGATGAAAAAAGAACGAA 6102
Db 1715 ----- 1715
QY 6103 ACAGGCGCGCTACAAAAACAGCCGAAAAAGTGAATAATCTTAACAACAGAAATCGAGCA 6162
Db 1716 -----ProProProProValGlnProGln----- 1723
QY 6163 CAAACAGAACCCGAGAGAGACACAACAAACAGAAACGAACATCAACAGCACACACA 6222
Db 1724 -----ProGlnAlaProThr----- 1728
QY 6223 ACAGAACTGACCTGGCGCAATGTAAAGCCATTTCTTCGAATTAACAGATGACAG 6282
Db 1729 ValThrValAspValCysSerLeuValLysThrLeuPheLys-----AspThrAsn 1745
QY 6283 GGTGAATAGAGGGTGTGATATCCAAAAAGCTATGACATAT-----CCTAAATGGCGT 6336
Db 1746 AsnPheSerAspAlaCysGlyLeuLys-----TyrGlyLysThrAlaProSerSerThrLys 1764
QY 6337 TGTATT-----GTAGTAACTTAAGAAAT 6363
Db 1765 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrGlyLysSerCysLysSerAsp 1784
QY 6364 GAAAAATGCGATATGATGCTGCTAGAGAGAAAAAATTTGATTAATTAATTAACATAT 6423
Db 1785 SerGlySerIleCysIleProProArgArgArgLeuTyrValGlyLysLeuGln--- 1803
QY 6424 TTAATTAATGAACCT-----GAAATTAAGGTGAC 6453
Db 1804 -----GluThrPalaThrAlaLeuProGlnGlyAlaAlaAlaProSerHisSerArgAla 1822
QY 6454 AATGATATTAAGAGCGCTTTATTAATGTCAGCAATAGAACTCAATTTTGTGTTA 6513
Db 1823 AspAspLeuArgAsnAlaPheLleGlnSerAlaAlaLleGlnThrPhePheLeuThrPasp 1842
QY 6514 AAATATATTAATGAAC--AATCGTCACAGAAATGA-- 6549
Db 1843 ArgTyrLysGluGluLysLysProGlnGlyLysPheSerGlnGlnAlaLeuSerGlnLeu 1862
QY 6550 -----TTGCAAAATGAGACA 6564
Db 1863 ThrSerThrTyrSerAspAspGluGluAspProProAspLysLeuLeuGlnAsnGlyLys 1882
QY 6565 ATTCAGATGAAATTTAAAGAAATATGATATATCATATGTGTATTAAGATGTTT 6624
Db 1883 IleProProAspPheLeuArgLeuPhePheThrLeuGlnLysPyrArgAspIleLeu 1902
QY 6625 TTGGAACGTATATTTCTAATGATTAATAAATAATTAATCTGATCAATATAGTAACAA 6684
Db 1903 ValHisGlyGlyAsnThrSerAspSerGlyAsnThrAsnGlnLysSerAsnAsnAsnIle 1922
QY 6685 ATTCATGATGAATTAATTAAGAAAAACAGATTAATAAAGATGAAGATTAACGTAAA 6744
Db 1923 ValLeuGlnAlaSerGlyAsnLysGluAspMetGlnLysIleGlnGlnLysIleGln 1942
QY 6745 ATA----- 6747
Db 1943 IleLeuProLysAsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAsp 1962
QY 6748 -----TTTGGGAGAAAAATTAATAATTTATTTGGCAAGAAATGATATGATTAACAT 6804
Db 1963 LysThrPyrAsnGlnHisAlaGlnSerLleThrLysGlyMetIleCysAlaLeuThrTyr 1982
QY 6805 -----CATCTCAGACGAGAAACGAAAAAGAAAAATTAACA 6840
Db 1983 ThrGluLysAsnProAspThrSerAlaArgGlyAspGluAsnLysIleGlnLysAspAsp 2002
QY 6841 GATAATTAC----- 6849
Db 2003 GluValTyrGluLysPhePheGlySerThrAlaAspLysHisLysIleThrAlaSerThrPro 2022
QY 6850 -----CAGTACAAAT--GACATGACCAAACTG----- 6873
Db 2023 ThrGlyThrTyrLysThrGlnTyrAspTyrGlyLysValLysLeuGlnLysPheSerGly 2042
QY 6874 -----ACGCTTCCTTGAAGAGCTTTGTAATAAGGCC 6906
```

Db 2043 AlaLysThrProSerLaserSerAspThrProLeuLeuSerAspHeValLeuArgPro 2062
QY 6907 CAATTTTGGAGAGTTCACAGAAATGGGAGAGAAATTTTGTAAATAGAGAGAGAACAG 6966
Db 2063 ProTyrPheArgTyrLeuGluGluTutPglGlyGlnAsnPheCysLysTyrSgTgShLys 2082
QY 6967 TTGTTAAATTTGGAGGGGGCTGTAAAGAAATATGAG- 7002
Db 2083 LeuAlaGlnIleLysShSglucYsLysValGluGluAsnGlyGlyLysSerArgArgGly 2102
QY 7003 - 7017
Db 2103 GlyIleThrArgGlnIleTyrSerGlyAspGlyGluAlaCysAsnGluLeuProLysAsn 2122
QY 7018 GAGCGTAAGACACAGAA- 7062
Db 2123 AspGlyThrValProAspLeuGluLysProSerCysAlaLysProCysSerSeryTyrArg 2142
QY 7063 AATTTTATTAAGAGTGGAAACTGAATATGAAAGACAAAGAAAGTTCAAAAGAGAT 7122
Db 2143 LysTrpIleGluSerLysGlyLysGluPheGluGluGlnGluLysAlaTyrGluGlnGln 2162
QY 7123 AAGAT- 7170
Db 2163 LysAspLysCysValAsnGlySerAsnLysShAsp- 2174
QY 7171 GAGAAAGCAACATGTCATGATATTAACATGAATTAAGAAATATAGTGGCAAT 7230
Db 2175 - 2175
QY 7231 AAGGATTTGTTCTGTATGCAAAAACCTTCTTACAACTCCAAAACACACAAACATCA 7290
Db 2176 GlyPheCysGluThrLeu- 2182
QY 7291 CAATTCATCGCATATGATATGACAGATGCGTGATTTATGTCGTAAGATTTAAAC 7350
Db 2183 ThrSerSerLysAlaLysAspPheLeuLysThrLeuGly- 2195
QY 7351 AAGTGTGAGTGCCTGCAACTTTCAAAAGAGATATGAT- 7392
Db 2196 - 7392
QY 7393 - 7440
Db 2214 PheLysH1ThrLysAsp- 2230
QY 7441 AAGCAGCATATATTTATCTAAAGAGAGAAATATATGATTTTACCTTGAAGAA 7500
Db 2230 - 2230
QY 7501 AATTTTATCTATGATGCTGTACAAAGAGAAAGAAATAGTTGACTATTAAT 7560
Db 2231 - 2238
QY 7561 AATCTTGGAGTCTTAAGAACTTATGACCTGATTAATATATAGAGAGAAAGAACTT 7620
Db 2239 - 2242
QY 7621 TGTGAATATAGAGAGAAATGCTTTAAGTATGATTAATGAATGAATATGAATTA 7680
Db 2243 CysArgAsn- 2247
QY 7681 TCAAACTTATCAGAGAGAAAGAGATGATGTACTCCAAAGAGACATATATGTC 7740
Db 2248 Ser- 2248
QY 7741 TTAAGAAATTTAGATGAATTAATATGAAAGCTTAAGATATGATATATCTCTTAA 7800
Db 2249 - 2265
QY 7801 ATGTTTCGTGAACATGACAGAAATGAAGAAATAGACATATATAAAACCTTCAACTCA 7857
Db 2266 ArgValSerAlaAspSerLysSer- 2278

QY 7858 - 7911
Db 2279 GlyLeuGluAsnAlaLys- 2284
QY 7912 GATCTGGGTGACATAGTATAGAGACAGATATGTTAGCAATTTGGTGTACTTACCTCC 7971
Db 2285 - 2288
QY 7972 GTAGAAATTAATATATAGATTTTGAATACATATATGCAAAATGAGAAATTAAT 8031
Db 2289 - 2293
QY 8032 AAGGTAGAAATTAATACACAGATGACAAAGCTTCTGCTTGTGGATGCTAAT 8091
Db 2293 - 2293
QY 8092 AGAAAGATATTTGGAAGCA- 8133
Db 2294 ArgLysAspGluTrpLysCysArgAsnValCysGlyTyrValValCysLys- 8132
QY 8134 GATGCAAACTTTTGAAGAGAGAAATGATGATTTGAACGCTAATTAATACAA 8193
Db 2313 AsnValAsnGlyGluAlaLysGlyLys- 2321
QY 8194 GATTAAGTGTGACATTAAGAGACATCCACCTGTGATTAATATATACCTCAACGTTTGA 8253
Db 2322 - 2328
QY 8254 TGGATGACTGAATGCTGTGAATATTTATGTAAGCAGTATGAGAGAAATTTG 8313
Db 2329 LeuValLysArgTrpValAlaLysTrpPhe- 2344
QY 8314 AAAAAATCATGATGATCATGTAACATCTGACAGATGACAGATGAAGAAATATATGATGA 8373
Db 2345 LysShS- 2355
QY 8374 AAGTGTGAACAGTGTAAAGAGATGTCAAGATATTAATTAATTTGTTTAAATGAA 8433
Db 2356 GluIleSerProCys- 2368
QY 8434 TCTCTTTGATATACATCAATTAATTAACAAAGATTTATGACACCAATATATACA 8493
Db 2369 - 2382
QY 8494 AAAATCTACTATGATCATGTTCAAAATTTGTACAAAAGTTGAAACTTTTAAAGT 8553
Db 2383 - 2385
QY 8554 GATGTTCTGTGAGAGCTTTTCTGAAATATCTTCATGAAGAAAGTAAGTGTGATTAAT 8613
Db 2386 Gln- 2387
QY 8614 AATTTTAAAGAAATGATGCTTCTCAATATATACAGAAATATGCTTTCAGAGAAACACA 8673
Db 2388 Lys- 2406
QY 8674 AAA- 8691
Db 2407 GlnIleThrAspAlaAsnAlaLysAsnLysValIleLysLeuSerLysPheGlyAsnSer 2426
QY 8692 TGCAGTTGTACACTACCTTCAAGAAATCCATTTGATTAATGCTCCAGCAAAACAA 8751
Db 2427 CysGlyCysSerAlaSerAla- 2438
QY 8752 GATGATGTGAAGATTTACAACTTTTACCTCTGCTCGAAGAAATGATTAATTAAT 8811
Db 2439 AsnGly- 2440
QY 8812 CTTGATTAATTTGAACCATACCTTGTCTTAATAGTTAGATGATATACAAAGGTGATG 8871
Db 2440 - 2440

```

QY 8872 ATTCTCCAGAAGAGACATTATGTACAGACCTATCATCATATATATATAGAAAA 8931
DB 2440 ----- 2440
QY 8932 GGTGATTAAGAAATTTTAAAAAACCCTTCTACTTCTGCTTCAGTCAAGACAAATTG 8991
DB 2440 ----- 2440
QY 8992 TTAGGTCAAAAAATATATAATCGAAGAGAGTGTGCTTTGAGCAATGAATATAGTAT 9051
DB 2440 ----- 2440
QY 9053 GCAGATTAATCCGATATATATTAAGAACTGATATGATGACACTTATATCTGAAAA 9111
DB 2441 ---GluTyrLysAspAlaIle-----AspCysMetLeuLysLysLeuLysAspLys 2456
QY 9112 ATTTAAAAAATTTTGAACAATCAATGACACCGAAAAATCGTAAAAACATGCGGGA 9171
DB 2457 Ile----- 2457
QY 9172 AATTAATAGAGTCAGATATGCGACCGCTATGTATGTGATATAAATGCTACTTCAAA 9231
DB 2457 ----- 2457
QY 9232 GTAACATTAGATGAAGATGGTGTCAATTACCAAGATGAAGAACTATCATGTTCTT 9291
DB 2458 -----GlyCys----- 2460
QY 9292 CGTTGTTAATGTGATGGCCAAAGACAGCATGTAAAGAAAGCAATGATGATTCGA 9351
DB 2461 -----GluLysLysHisHisGlnThrSer 2468
QY 9352 TTTAAACAAAAATGT-----CCTCGTTCAAGCAAGATATTTTGAAGCG 9396
DB 2469 ---AspThrGluCysSerAspThrProGlnProGlnThrLeuGlnAsp----- 2483
QY 9397 TCAGATTAATTAAGACAACTGATGTCAAGATGATTAAGAAATATATTAAGCTGAAT 9456
DB 2484 ---GluThrLeu-----AspAspAspIleGluThrGluGlnLysLys 2497
QY 9457 ATATGATTAATAAATATACATGCAAAATCAATTAATTAATAATTAAGCAATTAAGTCA 9516
DB 2498 AsnMetMetProLysIleCysGlnAsnVal-----LeuLysThrAlaGlnGlnLysGln 2516
QY 9517 TCTTCAGTAAATATAGACAAATTAACCATCTGAAGAAATGTTCACTCATATTAATAATCA 9576
DB 2517 ---GlyGlyCysValProAlaGlnLysSerGlnGlnProAla-----Ala 2530
QY 9577 AAAGATTCATGCGGCTTTGGAGTTAATGATTAATGAATGAATGTTACAGAACAAAA 9636
DB 2531 ThrAspSer----- 2533
QY 9637 AATTAATGAATAATGAATTCACAAAGAGTACTAAAAAATATATCTGTTATATTTT 9696
DB 2534 GlyLysGlnThrProGlnGlnThrProValLeuLys----- 2545
QY 9697 GTTGAAGATGAACAACAAAAATCATGTAAGTGAATTAATAAGCAAGACAGCA 9756
DB 2546 -----ProGlnGlnLysLys 2550
QY 9757 ACAAGTCTGCTTAAGCACTCTATTTCTTTACACCCCATGATGATCTTTCTATCAAGCA 9816
DB 2551 ValProGlnProPro-----ProProProProGlnGlnLysAlaProAla 2565
QY 9817 CTTTATATCTCAACACATCGAGTAGCACAAATATGATCCTAAAAAT----- 9861
DB 2566 Pro-----IleProGlnProGlnProProThrProProThrGlnLeu 2579
QY 9862 -----GATATATGAAGAGTAGATCTGTTGTTATTTGATCGCGCTTAGGT 9909
DB 2580 LeuAspAsnProHisValLeuThrAlaLeuValThrSerThrLeuAlaIleProSerValGly 2599
QY 9910 TTG-----ATAGCGCTTCATTTGATGAAGAAAAAATTAATCGTCTGTG--- 9954

```

```

DB 2600 IleGlyPheAlaThrPheThrTyrPheTyrLeuLysLysThrLysSerSerValGly 2619
QY 9955 GACGTGTCGCTATACATGAATATCCCGCAGAGAGTATGAAATCCATCGTTGAATCC 10014
DB 2620 AsnLeuPheGlnIleLeuGlnIleProLysSerAspTyrAspIleProThrLysLeuSer 2639
QY 10015 AAAAAAGTACATACATATATGAAGTGTCCATTAAGCAAAAAACATATATATATG 10074
DB 2640 ProAsnArgTyrIleProTyrThrSerGlyTyrArgGlyLysArgTyrIleTyrLeu 2659
QY 10075 GAAGAGATACT-----AGTGAGATGAAGATTAATATATGTGGACTTATCTTCC 10125
DB 2660 GluLysAspSerGlyThrAspSerGlyTyrThrAspHisTyr----- 2673
QY 10126 TCTGATTAATCTTCATCCGAAGTCAGATGAAGATGCAATGATTAATGATATATATGTA 10185
DB 2674 SerAspIleThrSerSerGlnSerGlnLysGlnLysGlnLysGlnLysGlnLysGlnLys 2693
QY 10186 CCAGGTAGTCTTAATATTAATAAATCATGATGATAGAGTACTGAACATCAAAAAAGGAT 10245
DB 2694 ProGlySerProLysTyrLysThrLeuIleGlnValIleLeuGlnProSerClyAsnAsn 2713
QY 10246 ATACCAAGTAT----- 10257
DB 2714 ThrThrAlaSerGlyAsnAsnThrThrAlaSerGlyAsnAsnThrThrAlaSerGlyLys 2733
QY 10258 GATACCAAGTAT-----AATGATACACCCAGTACCAATGATTT 10296
DB 2734 AsnThrProSerAspThrGlnAsnAspIleGlnAsnAspGlyIleProSerSerLysIle 2753
QY 10297 ATTGATGATGAATGAATGCAATGCAATGATTTTGTATCATATTTA----- 10347
DB 2754 ThrAspAsnGlnTyrPasnGlnLeuLysAspGlnPheIleSerGlnTyrLeuGlnSerGln 2773
QY 10348 CCAATTAACGAACCAATTAATTAATTAACAAAGTGCAGATATTTCAATGAATACAGAACCT 10407
DB 2774 ProAsnThrGlnPro-----AsnMetLeuGlyTyrAsnValAspAsnAsnThrHisPro 2791
QY 10408 AATCTTATATCTGATTAATCCGAGAAACCTTTATATATCATATCATGATGAG 10467
DB 2792 ThrThrSerHis-----HisAsnValGlnGlnLysProPheIleMetSerIleHisAspArg 2810
QY 10468 GATTATTAATCTGGGAAAGAAATTAAGTAT----- 10497
DB 2811 AsnLeuPheSerGlyGlnGlnTyrAsnTyrAspMetPheAsnSerGlyAsnAsnProIle 2830
QY 10498 AATATTAATATGAGTACTAAT-----ACTAATTAATGATATTCGAATG 10539
DB 2831 AsnIleSerAspSerThrAsnSerMetAspSerLeuThrSerAsnAsnHisSerProTyr 2850
QY 10540 AATGCTAAGAAATGATCTTATATGAGTATGATTAATTAATGATTCACATGATGCTGCTA 10599
DB 2851 AsnAspLysAsnAspLeuTyrSerGlyIleAspLeuIleAsnAspAlaLeu-----SerGlyA 2870
QY 10600 AACCTATTGATATATATGATGAGATATG 10628
DB 2870 snHisIleAspIleTyrAspGlnMetLeu 2879

```

```

RESULT 4
AAW00384
ID AAW00384 standard; Protein: 2913 AA.
AC AAW00384;
XX 21-FEB-1997 (first entry)
DE Plasmodium falci-parum erythrocyte membrane protein.
DE Plasmodium falci-parum; erythrocyte membrane protein; malaria;
KW detection; identification; treatment; prevention; parasite.
XX Plasmodium falci-parum MC type.
OS

```

Key	Location/Qualifiers
Domain	62..394
Domain	/label= Duffy binding ligand domain 1
Region	607..648
Domain	/note="Cysteine rich motif"
Region	839..1272
Domain	/label= Duffy binding ligand domain 2
Domain	1482..1527
Domain	/note="Cysteine rich motif"
Domain	1706..2005
Domain	/label= Duffy binding ligand 3
Domain	2102..2349
Region	/label= Duffy binding ligand 4
Region	2354..2398
Domain	/note="Cysteine rich motif"
Domain	2450..2475
Domain	/note="Putative transmembrane domain"

WO9633736-A1.
 31-OCT-1996.
 26-APR-1996; 96WO-US05798.
 27-APR-1995; 95US-0430908.
 (AEFY-) AEFYMAX TECHNOLOGIES NV.
 Baruch DI, Howard RJ, Pasloske BL;
 WPI; 1996-497376/49.
 DR N-PSDB; AAT41852.
 New Plasmodium falciparum erythrocyte membrane proteins - used to develop products for the diagnosis, treatment or prevention of malaria parasite infections
 Claim 1; Figure 12; 149p; English.
 A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte membrane protein 1 (PFEMP1) or active fragments or analogues of that protein can be used in the treatment or prevention of symptoms of a malaria parasite infection. The polypeptides can inhibit, block or reverse the sequestration of erythrocytes in patients suffering from malaria. Nucleic acids derived from the PFEMP1 gene can be used as probes and primers to identify a Plasmodium falciparum parasite, the primers used to generate characteristic amplification patterns from different P. falciparum strains. Antibodies specifically immunoreactive with the PFEMP1 polypeptide or its fragments may be used in diagnosis of malaria infection. This is the PFEMP1 protein of the MC type of Plasmodium falciparum. An alternative, truncated PFEMP1 protein is given in AA00385.

Alignment Scores:
 Pred. NO.: 9.53e-212 Length: 2913
 Score: 3128.50 Matches: 1003
 Percent Similarity: 38.47% Conservative: 454
 Best Local Similarity: 26.49% Mismatches: 958
 Query Match: 16.08% Indels: 1373
 DB: 17 Gaps: 149

US-10-087-013-1 (1-10628) x AA00384 (1-2913)
 97 AAAAGTGCAGAAATGTTTGGAAACGTTATGCCAAAAATATA----- 1388
 13 LysAspAlaValSHSLdaLeuAspArg1LeGluGluValTYLysGluYsValGlu 32
 139 AGACATGCATCAAAATATATGCAAAAGAACATCGATTCGTTGAAGAAGGGAGTTGACGAAA 198
 33 AsnaSpaLaGluYsTYLysYs-----AlaLeuYsGluYsAsnLeuGlu 48

OY	199	GCACAAATTCGGTGGTGGCTCTCTAGCCGACAGTAATAACCAATATTTATTTATTCACAT	258
Db	49	AlaLysGlyIleGlyIleuAlaSer-----	Prosn 60
OY	259	CCATGTATTTAGTCATTAAGCAACATACATTTTCGCTAATGATGATGTGATTTGACA	318
Db	61	ProGlyLysLeu---TyrGlnAspTyrTyrAlaAsnAlaGlyLeu-----ArgLysArg	77
OY	319	CATCTTCCCATGGTGTAGACAAACCGATTTGATGAAGATGAAAGATTCGATGTGGA	378
Db	78	TyrProGlyAlaAsnArgGlnThrValArgPheSerAspGlyTyrGlyGlyGlnCysThr	97
OY	379	---AATAAATACGTAATTTATAAAGAAAAAATGATGCGTAT---GCGTGGCCGACCT	432
Db	98	PheAsnArgIleLysAspSerGluAsnAsnAspAsnSerLeuGlyAlaCysAlaProTyr	117
OY	433	AGAGACGACATATGTGTGATAAAACCTTGGAACCTGTAAATGATTAATACCCAAAT	492
Db	118	ArgArgLeuHisLeuCysAspTyrAsnLeuGluLysMetGlyLysThrSerThrLys	137
OY	493	ATTCATGATTATTTGGAAATGTACTAGTATACGAAATACGAAGGTAAATCTGTT	552
Db	138	---HisGlyLeuLeuLeuAspValCysMetAlaAlaLysTyrGlnGlyAspSerIleLys	156
OY	553	AATATATC---CCACATTAAGGACCT-----TCAGACGCTGTACTGCT	594
Db	157	ThrHisTyrThrLysHisGlyIleuThrAsnProAspThrLysSerGlnLeuCysThrIle	176
OY	595	CTTGACGACGAAGTTTGCACATATATAGTGATTTGTAAAGCAATTCATATGTTT-----	648
Db	177	LeuAlaTyrSerPheAlaAspIleGlyAspIleValArgIleLysAspLeuTyrIleuGly	196
OY	649	-----AACCAAATGTCCATGACAAATGTGAAGACGGGTCCGACAGGTTTTC	696
Db	197	TyrAspAspLysGlnLysAspGlnAlaGlyLysLeuGlnAsnAsnLeuIleGlyIlePhe	216
OY	697	AAGAAATACATGTGATGGAAGAATGAATGAATGAATTCACATCCGATGATCT	756
Db	217	LysLysIleHisGlnAsnLeuGlnThrGlnAspAlaLysAspHisTyrLysLysAspLc	236
OY	757	GGAAATTTATTAATTAAGAAGACGATGGTGAATGCAATTAAGAATTAATATGGA	816
Db	237	GluAsnTyrTyrGlnLeuAlaGlnAspTyrThrAlaAsnAspSerThrValThrLys	256
OY	817	GCTATTAACATGTGATGCATCATTAATCTGATATTTTATTCATACGAAAGTATACA	876
Db	257	AlaIleHisCysHisAlaGlnLysIleSerAspLysTyrPheArgLysThrCysCysSerGly	276
OY	877	CCATATTTTCCAAATCT---AATGCGGCGATTAACCAAGGAAG-----	918
Db	277	GluThrThrAspAspLysCysArgCysLysAspGlnGlyLysAsnGlnThrAsnGln	296
OY	919	GTTCCCTACCAATTTAGATTTGTGCTCCCTCAATATTTACGTTGGTGGACGAGGAADA	978
Db	297	ValProThrTyrPheAspTyrValProGlnTyrLeuAlaGlyPheGlnGlnThrLysGln	316
OY	979	GAGTTTGCSSGAAAAAGAAATTAATTAATGAAAGAGTCAAGAGCTCTGTGTAATGAC	1038
Db	317	AspPheCysArgLysArgLysLysIleGlnAsnAlaIleLysAsnCysArgLysGln	336
OY	1039	AAA-----GAGCGCTAATTTGTAGTCAATTAAGGACATGATTTGAACGAACTATTGG	1092
Db	337	LysGlnLysGlnLysGly---TyrCysAspLeuAsnGlyTyrAsnCysGlnGlnThrAsnLys	355
OY	1093	AAAAAGGATTTTGCATTTGATTAATTAAGTACTAGCTGCTGTGCAATTCGAAAGTT	1152
Db	356	GlyLysGlnIlePheValLysGlnAspAspCysHisLysCysSerValAlaCysAspArg	375
OY	1153	TTTGAGATTGGTGTAGGATCAACAAGACGTTTAAAAACCAAAAGAAAATATGADA	1212
Db	376	PheValLysTyrIleAspAsnGlnAlaGlyGlnThrAspLysGlnLysLysLysTyrAsp	395

[illegible][illegible]

Db 823 ThrGln----- 824
QY 3349 TGCATATTAATATTCAGATAATACAAAGAAATTCATGACACAGCAAAATGCTGTAGT 3408
Db 825 -----AspGlyValLysProAlaSerGlnLysPheValLysValCys 839
QY 3409 AATAGTGTATGAGCTTCAGTACAGTCCCAAAATCATATACAGAGAAATGTTATTTGAA 3468
Db 840 SerLeu-----ValAspLysAlaLeu----- 846
QY 3469 TTTTTCGGAATTAATACCAAAATGTCGCAAAAGTAATAAAGTGTACAGTATGAT 3528
Db 847 -----LysGlyLysLeuAsp 851
QY 3529 GAAAGTCTGTCAAT-----GCTACTACACACACAGTATGAAATGTTTGAGCA 3576
Db 852 AspAlaCysThrLeuLysTyrGlyLysAspAsnThrThrGlnSerThr----- 869
QY 3577 TATCTCATGATACAGAAATTTTGATGATTCATGACACAAATGATTTTGTGATGAA 3636
Db 869 ----- 869
QY 3637 AAAAGTGTAGTAAGTAAACGAAATATGCTTTAGAGTAAACCAAGCAAGCATGAT 3696
Db 870 -----ThrLysPro----- 872
QY 3697 GGTGCGTGTGTGTTAAAGTGTGATCGAAACGACAGGCTAGATAAACGAAAAA 3756
Db 873 GlyAlaAlaGlyThrProSerGly----- 880
QY 3757 AAAGCGAAGAAAGATACGAAATGTAACAGTGAATATTAATAAGAAACGAT 3816
Db 881 -----LysAspThr----- 883
QY 3817 GGAAGAAACAAGTAGAAGATGTCATCAAAAAAGATAGATGCCGATTCG 3876
Db 883 ----- 883
QY 3877 CAATCGCAATATTAATTTAGTGAAGACCTGCTGTATGCCCTTACAGACAA 3936
Db 884 -----GlySerLeu-----CysValProProAlaGlyArg 893
QY 3937 AAGTATGCTACATTTCTG----- 3957
Db 894 LysLeuLysValGlyLysLeuHisAspTrpAlaGlyGlyLysThrGlnAlaLysSer 913
QY 3958 -----GCAATGATTAAT----- 3969
QY 914 GlnGlnThrSerGlyGlyLysThrProSerGlyAsnGlnSerProSerGlyLys 933
QY 3970 -----GAAATATAAAATTTACATCACAAATTAATTAATAAGAA 4008
Db 934 LeuProGlnGlyProThrProGlnThrThrLysGlnThrProGlnSerLeuLeuHis 953
QY 4009 GCTTTCATCAAACTGACAGACAGCAAACTTCTTCATGCTATTTATTAAGTAAG 4068
Db 954 AlaPheValSerProProAlaGlyLeuArgTrpPheLeuProTrpHisLysPheLysGlnGln 973
QY 4069 -----GAT 4071
Db 974 TyrPysAlaGlnHisGlyAlaGlyAlaThrGlyGlnGlnThrLeuLeuLeuAsp 993
QY 4072 GGTGAAGCAAAATGAACTC-----GATAAAGATTAAGAAAGCAAAATTCCTCCGCAATT 4128
Db 994 GlyGlyGlyLysGlnGlnThrProAspLysLeuLeuLysThrGlnHisLeuProProAspPhe 1013
QY 4129 TTGAGATCCATGTTCTACAAATTTGGAGATTATAGAGATTTTATTTGCA-----ACAGAT 4185
Db 1014 LeuAlaTrpGlnMetPheTyrThrLeuGlyAspTrpLysPheLeuValGlyAsnThrAsp 1033
QY 4186 ATA-----TCAAAAGCTAGTGTGAGGAGAAATTAAGTAAAGCAAAATGATTC 4236
Db 1034 IleValValHisThrSerGlyAsnLysGlnAspMetGlnIleMetGlnAlaIleGlnLys 1053

QY 4237 -----CTTTCAAAATGCTGACCAAAATCTCTAATGCAAAACA----- 4278
Db 1054 LysIleGlnGlnIleLeuProThrSerGlySerSerProSerProProArgValThrGln 1073
QY 4279 -----CGCAAGAAATGCTGACACAAACATCTACTGAGATA 4314
Db 1074 ThrGlnHisSerValGlnAsnProAlaGlyThrTrpTrpAsnGlnLysLysIle 1093
QY 4315 TGGGAAGCTATGCTATGCTACATGTA----- 4341
Db 1094 TrpGlnGlyMetValCysAlaLeuLeuThrTyrAsnThrAspThrProSerGlyThrAlaPro 1113
QY 4342 -----AAAATGGGCAAAAAA-----GATGATTTTACGCAAAAC----- 4377
Db 1114 ThrGlnIleGlnGlnValAlaGlyThrLysLeuTrpAspGlnAsnSerLysAsnProLysIle 1133
QY 4378 -----TACGTTACACAAACGTCAAATTTAGTAC----- 4407
Db 1134 ProGlnTyrLysTyrAspGlnValLysLeuAspAspThrSerAspAlaLysThrThrGly 1153
QY 4408 -----AAAGCACCACTTGGAGAAATTTGGCAAAAGCAAGCCCAAG 4446
Db 1154 SerProIleProSerGlyGlyLysIleThrProLeuThrAspPheIleSerAlaGlyPro 1173
QY 4447 TTTTACGATGCTTAAACGATGATGACAGCACTATGCTATGACAGCAAAATATTTG 4506
Db 1174 TyrPheArgTyrLeuGlnGlnLysTrpGlyGlnThrPheCysLysGlnArgLysLysAlaGly 1193
QY 4507 AAGATGTGCAAGAAATTAAGTCA----- 4533
Db 1194 GlnLysIleLysGlnGlnCysArgLysArgTrpGlyHisGlnHisCysSerGlyAsp 1213
QY 4534 -----AATGACCA-----TTGAAGTGT 4551
Db 1214 GlyTyrAspCysThrArgTrpAspAlaAspArgAsnAspLysPheValAspLeuAsnCys 1233
QY 4552 GATACAAATGTAATAGAATGCGAGACTACGATTAATATATATGAA-----AAAAA 4608
Db 1234 -----ArgAspCysHisIleGlnCysArgLysTyrArgLysTrpLysPheAsp 1252
QY 4609 GAGTGAATTCACAAAGATTAATTTACAGAGTGAACGCGCAAAAGATTCGATAGA 4668
Db 1253 GlnTyrHisLysGlnGlnLysLysTyrGlnGlyGlnThrAspLysLeuThrLysAspLys 1272
QY 4669 CAACACATGCTGAATGTTACAGACTATACGAAAGTGAACAGATTTACTTGAC 4728
Db 1272 ----- 1272
QY 4729 AGGAAATTTACTGTAGTGTGTGATTAAGCTGGAAGTCTGTGTGTAAGAAAT 4788
Db 1273 -----SerSerGlyLysAsp----- 1277
QY 4789 ATTCATTTGTAAGAAACAGGCTTACTATGATCCGACAAACATTTGGTGCACAAA 4848
Db 1278 -----AsnAsnCysCysLysAspIleGlnLysHisLysSerAlaAlaVal 1292
QY 4849 TTTATGAAATGACGCAAAATATACATTTTCAGTAAAGTAAAGTCAAGAGATTA 4908
Db 1293 PheLeuLysGln----- 1296
QY 4909 GTAAGAGGCAAAACACAGTGTATTAAGTGCACAAACAAAGCTCTAATACTAAT 4968
Db 1297 LeuLysHisCysLysAsnGlyGlnThrSer-----GlnAsnLysGly-----AsnGlnGlnLys 1314
QY 4969 AACTGAAAGATTTGACTGAAGATGCTTT-----TCTTTCGTCGACTACGATATAG 5024
Db 1315 GlnLeuAsnLysLeuAspPheAspLysIleProGlnThrPheSerProSerThrTyrCys 1334
QY 5025 TTTTCATGATGAGAGGCAATTAATACAGATTCAGAAAGTAAAGATGAAGATGGTTCG 5084
Db 1335 LysAlaCys----- 1337

QY 5085 AAAAGATGTCAGATGCGCGCAACGAGAGGTGACAAATTGGGTCAATACCAAGA 5144
Db 1338 ---ProVal1TyrglyValAsnCysAsnGly----- 1346
QY 5145 AAAAAGAAAAAGAAAAATTAAGATGCGATGCGCAGCAAAATTTCTATGAGTCCC 5204
Db 1347 AsnIysAsnTgglYAsnGlyGlyThrAsn---GlyCysThrThr----- 1359
QY 5205 GCGTTGATGTCATGAAATATAGTTTATGATTTAAGAGATTAATTTCTAGGTATGA 5264
Db 1360 -----AsnAsnGlu-ProG1 1364
QY 5265 TAATTGGAGATGAAAGAAAAAGACCGAGAGAAATTTGACAGAAATTTAAACAAAA 5324
Db 1364 uAsnIysGluAsnAsp----- 1369
QY 5325 TGGAAATCATGTTGGCAAGAAAGATGATAGTACAGAG-----AATCCGCTAG 5375
Db 1370 -----LysGlyAlaAlaSerThrIleSerIleLeuIleAsnAspGlyse 1384
QY 5376 TACTGCGCGAAATTTTTCGAAAGCAAAATTAAGATGTGTGGAACGCAATGATATG 5435
Db 1384 rThr----- 1385
QY 5436 CGGGTACAAACGTGTAGGATGATGAAATAGTGGAAATAGTCAAGAAATGATGAGA 5495
Db 1386 -----AsnGlyAlaThrAsnGlyThrThrGlyThr-----ThrAspGluTh 1399
QY 5496 TCTAAAAAAATGTGGTCTGTCTACCTCAGATGATGATATCTATG-----GGGAAAA 5549
Db 1399 rLeuIysGluCys-----SerAspIysTyrAlaPhePheIysGlyLeuAr 1414
QY 5550 TCGCGATGAGGTACTGCGCATCATGTTCTTCGATGTTGCCGAATGGGTCGAGATTT 5609
Db 1414 gIysGlnGluThrPheCysGlnIys-----LysTyrGlyValAsnG1 1428
QY 5610 TTCGAAACATTAAGAAAGAAATTTGAGAAATTTGTAGGCGCTGTATGATATATCTGG 5669
Db 1428 nCys-----AsnLeuThrAsnAlaValAsnAsp---Thry 1439
QY 5670 TGTGTATATGAAGATTAAGAAAGAAATGTACAGATGCGGTACCAATATTAATAAT 5729
Db 1439 rPheAspIysAsp-----IleValPheAsnGluTh 1449
QY 5730 TATTAGTGTGGAACCAAGATATGAATAAACAATCAAAATATGAGATTAATAA 5789
Db 1449 ePheGlnAlaGlyPheLeuArGlyPhe-----ValHisAspTyrAsnIleLeuIysH1 1466
QY 5790 CAAAATATATTCGAGACATCTGTGGCAAAAGATGACAGACGCTCGCGAATATTTAGA 5849
Db 1466 sIysIle-----As 1469
QY 5850 CAACAAATTAATAAAATTTGTGAAATTAAGAGATGATGATGATGATGATGATGATG 5909
Db 1469 pProcIysIleIysGluIysGluIysAspIysThr-----GluHisIysCysIle-- 1485
QY 5910 ACATGTGTCACACAGCATTAACGATGTGTAATAGTCAAAATATAGCCCGCATCTTGA 5969
Db 1486 -----AsnGlyCysAsnIle----- 1490
QY 5970 CGATGAACCAAAAGAGTTGAAGAAAGTGTATGTCAGAGCCACAGAGTCCACACAG 6029
Db 1491 -----LysCysGluCys----- 1494
QY 6030 TGTACGAGAGGAAACACCGCTACACAGGGGTACTGATATCAAAAGAGAGCGCATCGAA 6089
Db 1495 -ValArg----- 1496
QY 6090 AAAAGAGGAAACAGAGCGCGCTACAAACAGCGCAAAAGTGAATCTTAACAC 6149
Db 1496 ----- 1496
QY 6150 AGAAATGCGAGCAAAACAGAACCCGACAGACACACAAACAGAAACGAAATC 6209
Db 1496 ----- 1496

Db 1496 ----- 1496
QY 6210 AACAGCAACAAACAGAAATCTGACGTGGGCAAAATGTAAAGGCCATCTTTCGAATA 6269
Db 1496 ----- 1496
QY 6270 ACCAGATAGCAGGCGTGAATAGAGGCTGTATCCAAAAACGTATGCAATATCTTA 6329
Db 1497 -----Lys 1497
QY 6330 ATGGGCTTATATGATAGTATAGTATAGAAATGAAATGCAATATGATGATGATGATG 6389
Db 1497 strLeuGluIleIysGly-----AsnGluThrGly----- 1507
QY 6390 GAGAAAAAATATATGAT 6449
Db 1508 -----AsnIleIys-----LysHisTyrAsnIleAs 1516
QY 6450 TGCAATGATATTAAGAGAGCGCTTTATTAATGTCACCAATAGAAACTCAATTTTGTG 6509
Db 1516 nSerAsnAspIysGluThr----- 1523
QY 6510 GTTAAATATATATATTAAGAAATCTGCAGACAGAAATGAATTCGAAATGGAACATTC 6569
Db 1523 ----- 1523
QY 6570 ACATGATTTTAAAAAT 6626
Db 1524 -----IleAlaTyrAsnValIysSerTyrPheValAspGlnGlyLeuPh 1538
QY 6627 TCGAATGATATATCTAT 6686
Db 1538 eAspThrAspTyrIysIysAlaGlnIysVal----- 1549
QY 6687 TCTCATGAAATTAATAAGAAAAAAGAGATTAAGAAAAAGAGATTAAGATTAAT 6746
Db 1550 -----GluAspGluIysGluArgIys 1557
QY 6747 ATTTGGAGAAAAATTAATAATTTATTTGGGAAGAAATATATATATATATATATATAT 6806
Db 1557 sIleThr-----GlyCysThrGlyH1 1564
QY 6807 TCTCAGACAGAA---AACGAAAAAGAAAAATTAAGATATATATATATATATATATAT 6863
Db 1564 s-----AspGluCysSerGluIysGluIysGluIysAsnIys-----AsnPheI1 1579
QY 6864 GACCAACTGACGCGCTCCCTTGAAGAGTTGTAAAAAGCCCAATTTTGGAGATGTT 6923
Db 1579 eThrAsnLeuIleSerGluLeuGlnAspIysIleThrSerCysGln----- 1594
QY 6924 CACGAAATGGCAGAGAAATTTGTAATTAAGAGAGAGAAACAGTGTTAATAATGGAGC 6983
Db 1595 -----AsnIysHis----- 1597
QY 6984 GGGCTTAAGGAATATGATGATATGATAGTATAGTATAGCAGTATAGCAGAAATGTGAGA 7043
Db 1598 -----AsnProAsnGlyIysThr----- 1603
QY 7044 GGGCTGTATACATATCAAAATTTTATTAAGAAAGTGAAGAAATGATATGAAGACAAAG 7103
Db 1604 -AlaCys----- 1605
QY 7104 AGAAAGATTCAAAAAGATTAAGATGCAAAAGATTAAGATTAATCTTCTACTGAAG 7163
Db 1605 ----- 1605
QY 7164 AGACATAGAGAGGCAACAGTGTCTCATGAAATTTAAACATGAATAATTAAGAAATTA 7223
Db 1605 ----- 1605
QY 7224 TGGCAATAGGATTTCTTGTATGCAAAACCTTCTTCACTACCAAAACAAACACA 7283
Db 1605 ----- 1605

Db 1606 -----AspProPheProSerProThrPr 1613
QY 7284 ACAATCACATCATCGAGTCTATGATATAGCAGATCGCTGATATATGTCCTGAAGA 7343
Db 1613 oGluGluThrAspProLeuAspAspAspThrProAspProLeuAspAspAspGlnHis 1633
QY 7344 ATTTAAACAGTGTGAGTGTCTGAACTTCAAAAAAGGATCTATGATTCATCAAAAAA 7403
Db 1633 rGluGlnProLysPheCysPro----- 1640
QY 7404 AATTACTGAACCTTAATACCTTGTATGTGTAGAGAAAGCAGCATATTATTATCTAA 7463
Db 1641 -----ProProProProPheProMetThrCysValGluLysIleAlaLysGluLeuArg 1658
QY 7464 AGAAGCAGAAATATATGATATATCTGATGAGAGAAAA----- 7503
Db 1658 IGIuAlaGluGlyLysIleAsnAsnGluLeuLysGlyAsnGlyLysAspPheAsnGly 1678
QY 7504 -----TTTATACCTATTGAGCTTCAAAAG--GA 7529
Db 1678 scYsaAsnAsnValLysLysAsnGlyAlaValIleGlyGluGluSerCysLysPheG 1698
QY 7530 AAAGGAAGTAAATAGTATGACCTATATATCTTGATCCTTGAACCTTATATG 7589
Db 1698 uGlnThrTyrgLusAsnSerValAsnAsnIleAsn----- 1709
QY 7590 ACCTGATTAATATATAGAGAAAGAACCCCTGTGAATATAGAGAAATGCTTTAA 7649
Db 1710 -----AsnLysCysLysAspAsnGlnAsnGluArg 1721
QY 7650 GGTAGATTTGATGAGAAATGTTACAAAATTCAAAGTCTATCAGAG--AAAAAG 7706
Db 1721 sIleGlyGlnLysTrp-----AsnPheLysTyrgLysIleGlyThrIleArgLys 1737
QY 7707 AGTATGTATACCTCCAGAGAGACATATGCTTGAAGATTTGATGATTAAT 7766
Db 1737 pLeuCySIIeArgProArgTrgGluHisMetCysLeuAspAspLeuSerMetLeuGly 1757
QY 7767 TGAAGACCTTAAAGATAGTATATCTCTTAAATGTTGCTGCAAGTCAAGAAATGA 7826
Db 1757 gThrThrIleSerAspSerSerAlaLeuLeuLysIleGlnGluAlaLysSer 1777
QY 7827 AGCAATAGCATATATATAAACTTCAACTCAGAGAACGGGTGGCAATGATCAATATG 7886
Db 1777 uAGAspAspIleIleArgLysLeuLeuGlnGlnAsnSerCysAspGlnHisTrgIleC 1797
QY 7887 TGATACTGATTAATATAGTGTGCTGATCTGGGTGACATGATAGAGAAACATATGT 7946
Db 1797 sAspAlaMetLysTyrgSerPheAlaAspLeuGlyAspIleIleArgIlyArgAspLeu 1817
QY 7947 AGCAATGTGTACTTACTTACCTCCGTAGAAATATAATATATAGTTTGAATACAT 8006
Db 1817 pAsnLysAsnSerLysGlnLysGlyLeuGlnLysArgLeuGluLysAlaPheIleAsn 1837
QY 8007 ATATGGAATATGAGAAAT-----AAAAAT-----AAAGTAGAAATATAACACGA 8054
Db 1837 eTyrgAsnLysLeuGlnAsnAspLysAsnLysTyrgLysAspArgProLysTyrgLeu 1857
QY 8055 TGTACAAACCTTCTGCTGCTGCTGAGTGTATATAGAAAGATATTTGGAAGCAT 8114
Db 1857 nLeu-----ArgSerAspTrpTrpAspAlaAsnArgLysHisIleTrpAsnAla 1874
QY 8115 GACGTGCAGAACGACAGAGATCAAACTTTTAGAAAAGAGAGATGAT----- 8166
Db 1874 tThrCysAlaIleProAspAspAlaLysPheLeuLysAsnProAsnAspThrSer 1894
QY 8167 -----GGATTGAACGCTATACATTAATACAGATTAAGTGTGACATTAAGACGATCC 8219
Db 1894 ySerSerSerSerLysGlyIleMetThrThrHisSerAspCysGlyTyrgLysPheGlu 1914
QY 8220 ACCTGTGATGATATATATCTCAACGCTTTCGATGATGATGATGATGATGATGAT 8279
Db 1914 oProAspTyrgAspTrpIleProGlnProPheArgTrpMetGlnGluTrpSerGluSer 1934
QY 8280 TTGTAAAGCACTGATGAGAAATTTGAAAAATTTAAAAATCATGCTGACTGT---AA 8336
Db 1934 eCysLysLeuLeuAsnGlnGluMetGluGlnPheGluLysTrpCysGlyGluCysLys 1954
QY 8337 AACATCTGACAGATGCAAGAGATATATGATGAAATTAAGTGTGAACAGTGAACGAG 8396
Db 1954 sAsnSerIleThrCysGlyAspAspArgAsnGlyThrAsnCysGluAsnCysLysAsn 1974
QY 8397 AGTCAAGAAATATAAATTTTGTCTTAAAGCAATCTCATTCGATATACAAATCAA 8456
Db 1974 nCysGluLysTyrgLysLysLeuIleHisAsnTrpLysLeuGlyPheAsp----- 1990
QY 8457 TAAATCAAAAGATTTGATGAAACCAATATAT-----ACAAAATCTCTAC 8504
Db 1991 -LysTyrgLysGluIleTyrgAsnGlu---IleTyrgAsnAsnLysAspSerLysIleAsn 2009
QY 8505 TTATGATCATGTTCAAAATTTTGTACAAAAGTGAAGTAAAGTGAATG--TC 8561
Db 2009 rAsnGluTyrgPheLysPheLeuGlnLysLeu-----LysAspLysCysLysG 2026
QY 8562 TGTGAGAGCTTTTCTGAATATCTCATGAAACCAAGTAAAGTGTGATTAATTAAT 8621
Db 2026 uLeuAsnSerSerAspLysCysIleAspGluAlaThrHisCysThrLysTyrgLysPhe 2046
QY 8622 TGAATATGATGTCTTCCAAATATACGAACATATGCTTTCGAGAAACCAAAAGTAA 8681
Db 2046 r---AsnSerGluAsnLysAsnHisAsnGlyAlaPheLysAsnProProLysGlu 2065
QY 8682 TAAAGAACCTTGCACTGTGACACTACCTTCTAAAGATTCATGATGATATGTCCTACCA 8741
Db 2065 rGluLysAlaCysLysCysAspAlaPro-----AspProLeuAspAsnCysProLys 2083
QY 8742 TCAA-----AACAAAGATGATGTAAGGAATTAACAACCTTACTCTGCTCGAAGA 8795
Db 2083 pSerAlaThrTyrgLysAlaCysAsnThrLeuLeuProThrLysLeuGluSer 2103
QY 8796 TGATATGATTAATTAATCTGATTAATGCAACGATACCTTGTCTTAATAGTTCAGATGA 8855
Db 2103 sThrPheAsnAsnAspAspAspSerTrpAspThrSerPheAlaGlnThrSerProArg 2123
QY 8856 TAACAAAGGTGTATGATCTCTCCAGAGAAAGACATTTATGACAAAGCTATCCTG 8915
Db 2123 pAsnThrGlyValLeuValProProArgTrgArgGlnIleCysLeuLysAsnIleThr 2143
QY 8916 ATATAATTAGAAAAGGATTA---GAAATTTAAAAAAAACCTTACTCTGCTG 8972
Db 2143 r---LysLeuArgSerIleGluLysIleAspAspPheLysAlaGluLeuMetThr 2162
QY 8973 TTTCACTCAAGCACAATTTAGCTCAAAAATTAATTAATCGAAGAGAGTGTGCTTGA 9032
Db 2162 aTyrgAsnGluGlyLysLeuLeuGlnGlnLysLeuTyrgLysAspArgAspAlaThrLeu 2182
QY 9033 GCGAATGAATATAGTATGCAATATTCCTGATATATTAAGTAAGCAATGATGATGA 9092
Db 2182 nAlaMetLysTyrgSerPheTyrgAspTrpTrpGlyAspIleValLysGlyThrAspLeu 2202
QY 9093 CACTCATATCTGAAAAAATTAATAAATATTT-----GAAACATC 9134
Db 2202 rThrAlaProLeuAspLysLeuLysThrLysLeuAsnValLeuLeuLysGlyAspGly 2222
QY 9135 AAATGAAGCAACCGAAATCGTAAACATGCTGGGAAATTAATAGACGTCAGATATGGA 9194
Db 2222 rAsnGluIleLysGluAspArgGlyLysTrpTrpTrpGluAsnArgThrArgValTrp 2242
QY 9195 CGCTATGATATGATATTAATATCTCTACTTCAAAAGTAAACATTAATGATGAAGATG 9254
Db 2242 sAlaMetLysCysGlyTyrgLysAlaIleGlyLysIle-----GluGluArgAsp 2260
QY 9255 TCATATACCAAGATGAGAAACATATATGATGATGATGATGATGATGATGATGATG 9314
Db 2260 sSerLeuPro---AspAspAsnThrHisGlnPheLeuArgTrpPheArgGluTrpSer 2279

QY 9315 GCAAGCATGTAGAGAAAGAAAGCATGTAGTTCATTAATAAACAATGCTCGTTC 9374
 Db 2279 UHSPHCysAlaLysArgGlnLysLeuPheAsnGluValLysArgGlnLys -----AL 2297
 QY 9375 AAACGAGATTAATTTTGAAGCGTCAGAAATTAAGACACCGGAGTCAG----- 9426
 Db 2297 aSeAlaGlnCysLLeIleGluTyrGlyThrLLeAspProProValCysGluGluAlaC 2317
 QY 9427 -AATGATATTAAGAAATATATATAGCTTGAATATATGATATAAAATATGAGAAATCT 9485
 Db 2317 sThGlnTyrArgAspTyrLLeThrArgLysLLeGln----- 2329
 QY 9486 AAATATTAATTAAGCAATTAATAAGTCAATCTTACGTAATATAGCAATTAACCATC 9545
 Db 2330 -----GluTyrArgLeuLeuAsnTyrGlnTyrAsnThrAsnPheAsnLysLysAl 2347
 QY 9546 TGA--GAAATGTCAGCATATATATAAATGCAAAAGATTCATGCGCTTGAGGT 9602
 Db 2347 agluValThrLysAlaLProGluTyrPheAsnAspLys-----CysAsnAspLysC 2364
 QY 9603 AAATGATATTAATTAAGTATGTCAGGAAACAAATAATGAAATATGCAATTCAGAA 9662
 Db 2364 sAsnCysLeuSerLysTyrLLe-----AspLLeGluLys 2375
 QY 9663 AGTACTAATAAATTAATTCCTGTTATATTTTGTGAAGTGAACACCAAAATCA 9722
 Db 2375 slsTTPlysAsnMetClyAspSerPhe-----AspAspAsnAspLeuLysAsnLys 2392
 QY 9723 TGTACTAGATGAATAATATAAGACAGACAGTTCGCTTAAGACACTCTATTT 9782
 Db 2392 scysLLeCysArgGlnLLeLysProLysAsnPro-----ProLysLysValLysP 2409
 QY 9783 CTTTACACCCCATGTGATCTTTCTATACAGCACCTTATCTCAACACATCGATAC 9842
 Db 2409 ogLugluGlnHsThrProSerGluGlnAspThrProProLeuProProLysProAs 2429
 QY 9843 ACNATATGATCT-----AAAAATGATATTAAGAAAGTATCTTC 9884
 Db 2429 pAspLeuProProProAlaGluGluLProPheAsnArgAspLLeGluLysThrLLeP 2449
 QY 9885 TGTGTTATTTGATTCGCGTTAGTTCATGAGCGCTTCATTCATGAGAAATAATCA 9944
 Db 2449 opheGlyLLeAlaLLeuGlySerLLeAlaPheLeuPheLeuLysLysThrL 2469
 QY 9945 ATGCGTCTG-----GACTTGTTCGCTTACTGAATATCCGCAAGAGACTATGAGTCC 10001
 Db 2469 sSerSerValGlyAsnLeuPheGlnLLeuHsLLeProLysSerAspTyrAspLLeP 2489
 10002 TACGTTGATTCGCAAAATAGTATACATACATATAGAGTGTCCATTAAGGCAAAAC 10061
 2489 oThrLysLeuSerProAsnArgTyrLLeProTyrThrSerGlyTyrArgGlyLysArg 2509
 QY 10062 ATATATATATGAGAGGAGTACT-----AGTGAGAGTGAAGTAAATATATG 10112
 Db 2509 gTyrLLeTyrLLeuGluGlyAspSerGlyThrAspSerGlyTyrThrAspHsLysT 2527
 QY 10113 GGACTTATCTCTCGTGAATATCT-----TCATCGAAGTGTGATGAAGATTCGATAT 10169
 Db 2528 -----SerAspLLeThrSerSerSerGluSerGluTyrGluGluLeuAspL 2543
 QY 10170 TAATGATATATATGATACAGGATGCTCTAAATATATAAATGATGAGAGTACTGGA 10229
 Db 2543 eAsnAspLLeTyrValProGlySerProLysTyrLysThrLLeGluValLLeuG 2563
 QY 10230 ACCATCAAAAGGATATACCAAGTAT-----GATACACCAACT----- 10269
 Db 2563 uProSerLysAsnAsnThrThrAlaSerGlyLysAsnThrProSerAspThrGlnAsnAs 2583
 QY 10270 -----AATGATACACACAGTACGATGATTTATGATGATGATGATGATGAA 10322
 Db 2583 pLLeGlnAsnAspGlyLLeProSerSerLysLLeThrAspAsnGluTyrAsnThrLLeu 2603
 QY 10323 ACATGATTTTGTATCTCAATATTTA-----CCAAATACAGAACCAATATATATTA 10373

Db 2603 sAspGluPheLLeSerAsnMetLeuGlnAsnGluProAsnThrGluPro-----AsnMe 2621
 QY 10374 CAAAAGTCAGATATTCATCAATATACAGAACTTATATATTCGATTAATCTCGA 10433
 Db 2621 tLeuGlyTyrAsnValAspAsnAsnThrHisProThrThr-----SerArgHsAsnValG 2640
 QY 10434 AGAAAAACCTTTATATATCTATTCATGATAGGATTTATATCTGCGAAAGAAATTA 10493
 Db 2640 uGluLysProPheLLeMetSerLLeHisAspArgAspLeuTyrSerGlyGluGluTyrSe 2660
 QY 10494 TTATATATTAATATGAGTACTATATATATATATATATATATATATATATATATATAT 10553
 Db 2660 tTyrAsnValAsnMet-----ValAsnAsnAspLLeProLLeSerAlaArgAsnG 2677
 QY 10554 TTCTATATAGAGTAT 10613
 Db 2677 yAsnTyrSerGlyLLeAspLLeuLLeAsnAspSerLeu--AsnSerAsnLysValAspLLe 2696
 QY 10614 TATGATGAGATATG 10628
 Db 2697 tTyrAspGluLeuLeu 2701
 RESULT 5
 AAW22482
 ID AAW22482 standard; Protein: 2710 AA.
 XX AC AAW22482;
 AC 07-OCT-1997 (first entry)
 DT 07-OCT-1997 (first entry)
 DE Plasmodium Proj3.
 XX KW DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 XX OS Plasmodium falciparum.
 XX PN W09640766-A2.
 XX PD 19-DEC-1996.
 XX PF 07-JUN-1996; 96WO-US09508.
 XX PR 07-JUN-1995; 95US-0487826.
 XX RA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 XX PT Wellens TE;
 XX DR WPI: 1997-052231/05.
 XX N-PSDB: AAT72897.
 XX PS New malaria vaccines - contains cysteine-rich DBL family protein
 binding domains homologous domains of the Duffy and stalic acid
 binding proteins
 Disclosure; Page 50-56; 96pp; English.
 This sequence represents Proj3 of Plasmodium. Proj3 belongs to
 the Duffy binding like (DBL) family of genes which have homology to the
 Duffy antigen binding protein (DABP) and stalic acid binding protein
 (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 protein (DABP) are soluble proteins that appear in the culture
 supernatant after infected erythrocytes release merozoites. DABP and
 SABP mediate the binding of merozoites and schizonts to the erythrocyte
 surface. These proteins are necessary for erythrocyte invasion by the
 parasite. This sequence can be used in the compositions of the

QY 3634 GAAAA-----AGTATGATGAAGATAACAAAAATATGCTTTAGA 3675
Db 1298 LysLysHisIleGlyAlaThrSerThrThrLysGluAsnLysGlyTyrThrPheLys 1317
QY 3676 GATAAACCCACAGACCATGATGGCTGCTGCTGTGTTGTTAAAGTGATCGCAACCCACAGG 3735
Db 1318 GluProProProGluTyrAlaThrAlaCysAspCysIleAsnArgSerGlnThrGlu 1336
QY 3736 GTACAGATAAAGAAAAAGCGGAAAGAAAGATACGGAATGTAAACAGTGAAT 3795
Db 1337 -----GluProLysLysLysGluGluAsnValGluSerAlaCysLysIleValGlu 1353
QY 3796 GATATACCTTAAGAAAAACGATGGAAGAAACAAGTACGATTCATCCAAAAAGAAAT 3855
Db 1354 LysIleLeuGluGlyLysAsnGlyArgThrThrValGlyLysCysAsnProLysGluSer 1373
QY 3856 AGTAATGATATCCCATGTCGCAATGC---GCAAAATTAATTTAGTGGCAAGCCCTCGT 3912
Db 1374 -----TyrProAspTyrAspCysLysAsnAsnIleAspIleSerHisAsp---Gly 1389
QY 3913 GTGTGATGCCCTAGAACAAAAAGTTGCGTACATTTCTGGCAATGATTAATGA 3972
Db 1390 AlaCysMetProProArgArgGlnLysLeuGlyLeuTyrTyrIleAlaHisGluSerGln 1409
QY 3973 ATAAAAAATTACATCACACAGTTAATTTAAAGAGCTTTCATCAATCTGCAGCAGCA 4032
Db 1410 ThrGluAsnIleLysThrAspAsnLeuLysAspAlaPheIleLysThrAlaAlaAla 1429
QY 4033 GAACACNTCTCTCATCGCTTTATTTAAAGATAG---GATGCTAGAGAAATGAACTC 4089
Db 1430 GluThrPheLeuSerThrGlnTyrTyrLysSerLysAsnAspSerGlnAlaLysIleLeu 1449
QY 4090 GATAAAGAATTAAGAGCAAAATTCCTCCGCAATTTTGAGATCCATGCTTACACA 4149
Db 1450 AspArg-----GlyLeuIleProSerGlnPheLeuArgSerMetMetTyrThr 1465
QY 4150 TTTGGACATTATAGACATTTTATTTTGAACAGATATATCAAAAGCTCATGCTAGAGGA 4209
Db 1466 PheGlyAspTyrArgAspIleCysLeuAsnThrAspIleSerLysGluAsnAspVal 1485
QY 4210 AGTAACCTAAGAGCAATAGATTCCTTTTCAAAAATGTGACCAAAAATTCCTCAAT 4269
Db 1486 AlaLysLysLysAspLysIleGlyLysPhePheSerLysAspLysSerProSer 1505
QY 4270 GGAATAACACGACAGATGCTGACAGACATATGATGATGAGTATGAGTATGCTA 4329
Db 1506 GlyLeuSerArgGlnGluTyrPtyrLysThrAsnGlyProGluIleTyrPtyrGlyMetLeu 1525
QY 4330 TGTGCACCTACTAAAAATTTGGGCAAAAAAGATGAT-----TTTACCGAAAACTAC 4380
Db 1526 CysAlaLeuThrLysTyrValThrAspThrAspAsnLysArgLysIleLysAsnAspTyr 1545
QY 4381 GGTTACACACACGCAATTTAGTACAAA---AGCACCACTTTGGAGGAATTTGCCAAA 4437
Db 1546 SerTyrAspLysValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnGluPheAlaAla 1565
QY 4438 CGACCCAGATTTTACGATGCTTACCGATGATGATGATGATGATGATGATGATGATGAT 4497
Db 1566 LysProGlnPheLeuArgGlyMetIleGlyTyrGlyLeuGluPheCysAlaGluArgGln 1585
QY 4498 AAATATTTGAGAGTGTGCAAGAAAAATGTAACTCA---AATGACCAATTTGAAGTGTAT 4554
Db 1586 LysLysGluAsnIleLysAspAlaCysAsnGluLeuLeuSerThrGlnGlnCysAsn 1605
QY 4555 ACA-----GAATGATATAGCAATTCGACAGACTACCTTAATATATAGAAA---AAA 4602
Db 1606 AspAlaLysHisArgCysAsnGlnAlaCysArgAlaTyrGlnGluTyrValGluAsnLys 1625
QY 4603 AAAAAAGATGATTCACACAGATTAATATATTAAGGATGAAGCGCAAAAAAAGATTC 4662
Db 1626 LysLysGluThrSerGlyGlnThrAsnAsnPheValLeuLysAlaAsnValGlnProGln 1645

QY 4663 GATAGACAAACACATTTGCTGATGTTTACAGACTATATCTGGAACGAATGCACATTAAC 4722
Db 1646 AspProGluTyrLysGlyTyrGluTyrLysAsp----- 1656
QY 4723 TTGACACGAAATTTACTGTAAGTTGTTGATGAACCCCTGGAAGTGCCTGTGATCA 4782
Db 1657 -----GlyValGlnPro-----IleGln 1662
QY 4783 AGAAATATACATTTGTTAGAAAAACAGGCTTACTATGATGCCGACAAACATTTGGGTGC 4842
Db 1663 GlyAsnGluTyrLeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys 1678
QY 4843 ACAAAATTTATTTGAAAAATGACGACAAATATACTAACATTTGAGTAAGATGTCAAA 4902
Db 1678 ----- 1678
QY 4903 GGATTTAGTAAGAGGCAACACAGTGCCTATTAAGTGGCAAAACAAAGTCTTAATAC 4962
Db 1678 ----- 1678
QY 4963 TACAAATACCTGAAGAAATGACTGAAGATGCTTTTCTTCTGTCGACTACGTATA 5022
Db 1678 ----- 1678
QY 5023 TGTTCATGCAATTTGATGCGCAATTTATACAGATCAGAAAGTAAAGATGAATGGCTTG 5082
Db 1679 -----MetAspGlyAsnValLeuSerValSerProLysGluLys----- 1691
QY 5083 CGAAAAAAGATTTGATGGAAGAGGGGCGCACAGGAGGTACATTTGGGTCAATCTACAA 5142
Db 1692 -----ProPheGlyLysTyr----- 1696
QY 5143 GAAAAAAGAAAAAGAAATTAATAAAAGTCGATGCGCACAAATATCTTATGAGATC 5202
Db 1697 -----AlaHisLysTyrPro----- 1701
QY 5203 CCGCCTTTAGTCTATGAATATAGTTTATGATTTAAGAGATTAATTTAGGTATTT 5262
Db 1701 ----- 1701
QY 5263 GATAATTTGGAAGATGAAAAAAGAACCGAGGAAATTTGAAAGAAATTTTAACAA 5322
Db 1701 ----- 1701
QY 5323 AATGAAACATCACTTTGCAAGAGAGTATGATCTACAGAAATCCCGTACTACTGCG 5382
Db 1701 ----- 1701
QY 5383 CGAAAAATTTTCTGAAACGAAATTAAGAAATGTGTGTGAAACGCAATGATATCGGGTAC 5442
Db 1701 ----- 1701
QY 5443 AAACGTGTAGGATGATGAAATAGTGAATATAGTGAAGAAAGTATGAAGATCTAAAA 5502
Db 1701 ----- 1701
QY 5503 AAATGTGTTCTGTACTTCAGATGATGATATCTATGAGGAAAAATCGCGATGAAGT 5562
Db 1701 ----- 1701
QY 5563 ACTGCGTATCAGTTTCTTCGATGCTTTGCCGAATGGGCTGAGATTTTGCACAAATAAA 5622
Db 1701 ----- 1701
QY 5623 GAAAGGAATTTGAGAAATTTGTAGGGCGCTGTAAATGATTAATCTGTGTGATTAATGA 5682
Db 1701 ----- 1701
QY 5683 GATAAAGAAAGAAATGTACAGATGCGGTACACAAATATTAATAATTTATAGTGTG 5742
Db 1702 -----GluLysCys----- 1704
QY 5743 AAACACAGTATGAAAAAACAATCAAAAAATATGTTGAGATTAAGACAAATATATTC 5802

Pred. No.: 9.2e-201 Length: 2710
 Score: 2972.50 Matches: 865
 Percent Similarity: 40.06% Conservative: 442
 Best Local Similarity: 26.51% Mismatches: 873
 Query Match: 15.28% Indels: 1093
 DB: 21 Gaps: 120
 US-10-087-013-1 (1-10628) x AAY77904 (1-2710)

```

QY 25 ATTTAAATAATGGGATGACATCATCTTGAAGGAGATGCTAAAGCCCTATTATA 84
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 6 LeuAlaLysMetGlyProLysGluAlaIleGlyLysAspIleGluAsp----- 22
QY 85 AAAGAAGTCACAAAGATGCAAGAAATGTTTGAGACGTTATGCCAAATATAGACAT 144
   ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db 23 -----GluSerAlaLysHisMetPheAspArgIleGlyLysAspValTyrAsp 38
QY 145 CCATCAAAA--TATGCAAAAGAACATGTGATTCCTGTAAGAGGATTTGAGAAAGCA 201
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 39 LysValLysGluGluAlaLysGluArgGlyLysGlyLeuGlnIleArgLeuSerGluAla 58
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
202 GAATTCGTCGTGTCCTCTACGCCCAATAGCATATTATTATTCATATCCCA 261
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 LysPheGluLysAsnGluSerAspProGlnThrProGluAsp-----Pro 73
QY 262 TGTAAATTTAGATCATAGAACATCTACTAATTTACGGTATGATGTGATTTGAGACAT 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 74 CysAspLeuAspHisLysTyrHisThrAsnVal-----ThrHisAsnValIleAsn 90
QY 322 CCTTCGATGATGAGAACCAAAACCGATTGTGATGATGAGAACAACTGAAATG--GGA 378
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 91 ProCysAlaAspArgSerAspValArgPheSerAspIleGlyLysGlnCysThrHis 110
QY 379 AATTAATAATCGTAATTAATAAAGAAAAATGATGCTATAGCTGTGCGCCACTAGAGAA 438
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 111 AsnArgIleLysAspSerGlnGlnIleGlyAspAsnLysGlyAlaCysAlaProTyrArgArg 130
QY 439 CGCATATGTGTGATTAATAAAGTGAAGCTCTAATGATATTAATATCCCAAAATATTCAT 498
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 LeuHisValLysAspGlnAsnLeuGlnIleGluProIleLysIleThrAsnThrHis 150
QY 499 GATTTATGGGAATTTACTAGTTACAGCAAAATACAGAGTGAATGTTGTTAATAAT 558
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 151 AsnLeuLeuValAspValLysMetAlaAlaLysPheGluGlnIleSerIleThrGlnAsp 170
QY 559 CATCCA--CAATAAGGAACCT-----TCAGACGCTTGACTGCTTGCA 600
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 171 TyrProLysTyrGlnAlaThrTyrGlyAspSerProSerGlnIleCysThrMetLeuAla 190
QY 601 CCAAGTTTGCAGATATAGTATGATTTGTAAGAGAAATGATATGTTAAACCAATGTC 660
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 ArgSerPheAlaAspIleGlyAspIleValArgGlyArgAspLeuTyrLeuGlnLysAsnPro 210
QY 661 CATGAC-----AAAGTAGAAACGGGCTCCGAGAGGTTTTCAGAAATA 705
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 GlnGluIleLysGlnArgGlnIleGlnLysAsnAsnLeuLysThrIlePheIleLysIle 230
QY 706 CATGAT-----CGAATGGAACATGAAGTAAATAATGATTACATTCGATGATCT 756
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 231 TyrGluLysLeuAsnGlyAlaGluAlaArgTyrGlyAsnAsp-----ProGlu----- 246
QY 757 GGAATTTATTAATTAAGAACATGATGATGTAAGATGTAAGATTAAGATGAGAA 816
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 -----PhePheLysLeuArgGlnAspTyrPheThrAlaAsnArgGluThrValTyrLys 264
QY 817 GCTATACATGTGATGATCATATTAATCTGATATTTTATGCAATGCAAAAGTATACA 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 265 AlaIleThrCysAsnAla--TyrGlyAsnThrTyrPhe--HisAlaIleCysAsnArg 282
QY 877 CCATATATTTTCAATCTCAATATGCGGCATTAACAGGAAGGTTCCATCAATTTAGAT 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 283 GlyLysArgThrLysGlyTyrCysArgCysAsnAspAspGlnValProThrTyrPheAsp 302
  
```

```

QY 937 TATGCTCCATATATTTAGCTTGTCAGCAATGGGAGAAAGCTTTGCCGAAAAGCA 996
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 303 TyrValProGlnIleuThrLeuThrPheGluGluTyrPalaGluAspPheCysArgLysLys 322
QY 997 AATATTAAATTTGAAAAAGCTCAAGACCTCTGCTGCT-----AATGACAAAGAA 1044
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 323 AsnLysLysIleLysAspValLysArgAsnCysArgGlyLysAspLysGluAspLysAsp 342
QY 1045 CGCTTATATTTAGTCATATATGACATGATGTTAGCACACTATTTGAAAAAGGTATT 1104
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 343 Arg--TyrCysSerArgAsnGlyTyrAspCysGluLysThrLysArgAlaIleGlyLys 361
QY 1105 TTGCATTTGATTAATTAAGTGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1164
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 362 LeuArgTyrGlyLysGlnCysIleSerCysLeuTyrAlaCysAsnProTyrValAspTyr 381
QY 1165 TTAGGAGTCAACAGAACATTTAAAAAAGAAAAAAGAAATATGAAAGAAATACA 1224
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 382 IleAsnAsnGlnLysGluGlnIlePheAspLysGlnLysLysTyrAspGluIleLys 401
QY 1225 TCATAT-----TTATGCAACGATTAACAAATTTGTCAT 1257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 402 LysTyrGluAsnGlyAlaSerGlyLysSerArgGlnLysArgAspAlaGlyLysThrThr 421
QY 1258 AATATTAAATAGTGA--TATTATTAACAAATTTATGAAAACTTAAGAAACGAAATAT 1314
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 422 ThrThrAsnTyrAspGlyTyrGluLysLysPheTyrAspGluLeuAsnLysSerGluTyr 441
QY 1315 GCACTAATGACACTTTTAAATTTAAATTTAAATGAGAAAGATATGCT----- 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 ArgThrValAspLysPheLeuGluLysLeuSerAsnGlnIleCysThrLysValLys 461
QY 1363 -----AAAGCAGATTCACGAGAGAAAGATATTTACTTTTACTTAACAGTCT----- 1410
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 462 AspGluGluGlyGlyThrIleAspPheLysAsnValAsnSerAspSerThrSerGlyAla 481
QY 1411 -----GATGCAAAAGGATATTTATGCTGCAATATTTGCCAAGTGT 1455
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 482 SerGlyThrAsnValGluSerGlnGlyThrPheTyrArgSerLysTyrCysGlnProCys 501
QY 1456 CCCGACTCGGGGCTCAAAATGATGATGATTAATAATACACAC-----AATCAGATATGAT 1512
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 502 ProTyrCysGlyValLys-----LysValAsnAsnGlyLysSerSerAsnGlu 517
QY 1513 CGTGAACGTGAATAAT-----GAGACCTATTAACCTCCATGGGGT 1554
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 518 TrpGluGluLysAsnAsnGlyLysCysLysSerGlyLysLeuTyrGluProLysProAsp 537
QY 1555 GTGAAGCTTAAATATCATCGTCTTATATAGTATGATGTAAGCAAGGTGATATACAA 1614
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 LysGluGluTyrThrThrIleThrIleLeuLysSerGlyLysGlyHisAspAspIleGluGlu 557
QY 1615 AAATAGAAATTTTGTAAAC-----AGCTCA 1641
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 558 LysLeuAsnLysPheCysAspGluLysAsnGlyAspThrIleAsnSerGlyLysGly 577
QY 1642 ACTAATTACAAAGATTAATAATAT-----CAAAATGGGAATGC 1680
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 578 ThrGlyLysSerGlyLysGlyAsnSerGlyArgGlnIleuLysGlnIleuTyrLysCys 597
QY 1681 TATTAATAGATGAAATATTAATATAGATGTAAGTGAACAAATACTGAAATCAATAT 1740
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 598 ---TyrLysGlyLysAspValValLysValGlyHisAspGlnAspAspGluLysAspTyr 616
QY 1741 GATAATCCTAAG-----ATAATA----- 1758
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 617 GluAsnValLysAsnAlaGlyLysLeuCysIleLeuLysAsnGlnLysLysAsnLysGlu 636
QY 1759 -----TCATTTCAATATTTT 1773
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 637 GluLysGlyAsnThrSerGluLysGluProAspGluIleGlnLysThrPheAsnProPhe 656
QY 1774 TTGAAATTATGGGTTACATATTTATTAAAGGATTAAGTGAATGACAAACTTAA 1833
  
```

Db 657 PheTYTTRTPVAlahshmetleuLysAspSerIleHisThrLysLysLysLeuGln 676
QY 1834 ACTGTATA---AATAATCAACACCGCATTCG---ATTGATGATGTATGACAAATTCG 1887
Db 677 ArgCYLeuGlnsnclLysasnArgIleLysCysGlyAsnAsnLysCysAsnAspCys 696
QY 1888 TTATGTTTTCACAGATGGCTTAAACAAAGAAAGAAATGAAATGATTAAGAAACAG 1947
Db 697 GlnCysPheLysArgTrpIleThrGlnLysAspGlnTrpGlyLysIleValGlnHis 716
QY 1948 TTCACAAAAAAGAAATATACAG----- 1971
Db 717 Phe---LysThrGlnAsnIleLysGlyArgGlySerAspAsnThrAlaGlnLeuIle 735
QY 1972 -----CAATCGTATATGATTAATTAATCTTTTGAGGCTATATTT----- 2016
Db 736 ProPheAspHisAspTyrValLeuGlnTyrAsnLeuGlnGlnLeuPheLeuLysGlyAsp 755
QY 2017 -----TTTAAAGTTATGATTAATCACTTACAAAGATGAAAGCAAAATGCG 2058
Db 756 SerGlnAspAlaSerGlnGlnLysSerGlnAsnSerLeuAspAlaGlnIleAla----- 773
QY 2059 AAGCAACTATGCAAAATATATAAAGAAAAAATGAGTTTCCATTTGCAAAATAT 2118
Db 774 -----GlnGlnLeuLysHisLeuArgGlnIleIleGlnSerGlnAspAsnAsn 789
QY 2119 AGGGAC-----TATTAGACATGATCATAGACCTTGTTA 2154
Db 790 GlnGlnAlaSerValIleLysGlyValIleThrGlnLysAsnIleMetAspLysLeuLeu 809
QY 2155 GATCACTTAAAGCAAACTCCACGATATGT----- 2184
Db 810 AsnTyrGlnLysAspGlnAlaAspLeuLysLeuGlnIleHisGlnAspGlnGlnGln 829
QY 2185 -----AAGACAAATATACAAAGCAAGATGTGAACATCCCATATGCAACAAACCG 2241
Db 830 LysGlnLysGlnLysAspLysGlnLysIleGlnGlnGlnLysAsnThrGlyAsnPro 849
QY 2242 TGTGTTAACTCTCGTGAAGCAAGCAAA---CCCACTAAATATTAAGCAATGACAA 2298
Db 850 CysSerGlnGlnLysSerGlnLysAsnArgTyrProValLeuAlaAsnLysValAlaTyrGln 869
QY 2299 TACTTTAAAGAGTCATACAGCAAGCAAGCAAAATCGTCTTCAATAATGAAAGCA 2358
Db 870 MethHisIleLysAlaLysThrGlnLeuAlaSerGlnAlaGlyArgSerAlaLeuArgGly 889
QY 2359 AAGGCACAGCAAGCTATATATAACGTGGCGGTAGAGAAAGCAAGCTTCAAGCAATTTA 2418
Db 890 AspIleSerLeuAlaGlnPheLysAsnGlyArgAsnGlySerThrLeuLysGlnIle 909
QY 2419 TGTAGATATATGATTAACATTTATCTGTAATCTTGTTTCAATGACCAATGAT 2478
Db 910 CysLysIleLeuAsnLysAsnTyrSerAsnAspSerArgGlyAsnSerGlnGlyProCysThr 929
QY 2479 GGCAGAAAGCAGCGTATGATTAACAACAAGATTTGCTGTAAGCAAGTGAAGAGTG 2538
Db 930 GlnLys---AspGlnLysPheIleGlnValArgMetArgIleGlyThrGlnTyrSerAsn 948
QY 2539 GATCGCGAACAATGCGTAAAGATCAAGCAAGATGATTAATGCTCTGCAAGAACACAT 2598
Db 949 IleGlnGlnLysLysGlnIleThrSerTyrLysAsnValPheLeuProAlaArgGlnHis 968
QY 2599 ATATGTACATTCATTTGGAACTTTTACAAAGCAAGATGATCAACCACTTAATGATAT 2658
Db 969 MetCysThrSerLysLeuGlnAsnLeuAspVal-----GlySerVal 982
QY 2659 GTT---GATGATTTACTTAATTAATCTTTTGGGAGATCTTCTATACGCAAAATAT 2715
Db 983 ThrLysAsnAspLysAlaSerHisSerLeuGlnIleAspValGlnLeuAlaIleLysThr 1002
QY 2716 GAGCAACAACAGATATACGAATGTATTAAGAAAGAAATTAACCTTAAGGCGCCCAAGAA 2775
Db 2716 GAGCAACAACAGATATACGAATGTATTAAGAAAGAAATTAACCTTAAGGCGCCCAAGAA 2775

Db 1003 AspAlaIleGlnIleIleLysArgTyrLysAspGlnAsnAsnIle-----Gln 1018
QY 2776 GTAACAGACCA-----AAACACAGACAAATCATCTGTCAGCTATGATTCAGT 2826
Db 1019 LeuThrAspProIleGlnGlnLysAspGlnGlnIleAlaMetCysArgAlaValArgTyrSer 1038
QY 2827 TTTGCAGATATAGCTGATATTAATTCGAGAAAGATCTCTGGAAAGAAAGCGTGACATG 2886
Db 1039 PheAlaAspLeuGlnLysPheIleIleArgGlyArgAspMetTrpAspLysSerSer 1058
QY 2887 GTAACCTGCAGACATTTGCAAACTGTTTGTATATATCATATAGCTCAAAAGCG 2946
Db 1059 ThrAspMetGlnThrArgLeuIleThrValPheLysAsnIleLysGlnLysHisAspGly 1078
QY 2947 -----AAGCAAAATGATTAATATATATGATGAT-----GCCGCCAAATATTAATG 2994
Db 1079 IleLysAspAsnProLysTyrThrGlnLysAspGlnLysLysProAlaTyrLysLysLeu 1098
QY 2995 AGGCAAAATGCTGGCAACCTAATAGACCAAGATATGCGAAGCCATGAAATGTGATATA 3054
Db 1099 ArgAlaAspTrpTrpGlnAlaAsnArgHisGlnValIleTrpArgAlaMetLysCys----- 1116
QY 3055 AATATTTGAGATTAATCGGACACCAATCAACCAAGATATGCTATGCGGATTAAT 3114
Db 1117 -----AlaThrLysGlnIleIleLys----- 1123
QY 3115 GATCATACCAATGATGATATATATCCCAAAATTAAGATGATGACCGAATGGCA 3174
Db 1124 ProGlnMetProValAspAspTyrIleProGlnArgLeuAlaArgTrpMetThrGlnTrpAla 1143
QY 3175 GAATGCTACTGCAAGGTGCAAAAAAGAGATGATTAATGATGAGAGAGAGAGAGAG 3234
Db 1144 GlnTrpTyrCysLysAlaGlnSerGlnGlnLysAspLysLeuLysLysIleCysAlaAsp 1163
QY 3235 TGTAAAGATTAAGATTAATGTCAGCTGACCAAAAGAGAGTGTACAGTGTACAGAG 3294
Db 1164 CysMetSerLysGlnLysAspGlyLys---CysThr---GlnLysAspValAspCysGlyLys 1181
QY 3295 TGCACAGAAAGCTTGTATGATTAATGATTAATGATTAATGATTAATGATTAATGAT 3354
Db 1182 CysLysAlaIleCysAspAspTyrLysGlnGlnIleGlnLysTrpAsnGlnGlnTrpArg 1201
QY 3355 ATATATTCAGATTAATTAACAAGATTAATGATTAATGATTAATGATTAATGATTAAT 3414
Db 1202 LysIleSerAspLysTyrAsnLeuLeuTyrLeuGlnAlaIleLysThrThrSerThrAsnPro 1221
QY 3415 GGTATTTGAAGCTTCAGTCTGCAAAATCATATGAC-----AGGAATGTTAT 3465
Db 1222 Gln-----ArgThrValLeuGlnLysAspAspAspProAspTyrGlnGlnMetVal 1237
QY 3466 GAATTTTGTGGAATTAATACCAAAATGCGGCAAA-----AGTAAATAA 3513
Db 1238 AspPheLeuThrProIleHisLysAlaSerIleAlaAlaArgValLeuValLysAlaGln 1257
QY 3514 AGTGTACTAGTGAAGAGTGTGATGATGATTAATGATTAATGATTAATGATTAATGAT 3573
Db 1258 AlaGlySerProThrGlnIleAlaIleAlaIleAlaProIleThrProTyrSerThrAla 1277
QY 3574 GCATATTCATGATACAGAAATTTGATGATGATGATGATGATGATGATGATGATGAT 3633
Db 1278 GlnTyrIleHisGlnGlnIleGlnIleGlnIleGlnIleGlnIleGlnIleGlnIle 1297
QY 3634 GAAAAA-----AGTATGATGATGATGATGATGATGATGATGATGATGATGAT 3675
Db 1298 LysLysHisGlnAlaThrSerThrThrThrLysGlnLysAsnLysLysLysLysLysLys 1317
QY 3676 GATTAACCAACAGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3735
Db 1318 GlnProProGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1336
QY 3736 GTACAGATTAATAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGCAAA 3795
Db 1337 -----GlnProLysLysLysGlnGlnLysAsnValGlnLysLysLysLysLysLys 1353

OY	3796	GTTTACTTAAGAAACGATGGAACAACAACTAGACAGATTGTGCATCCAAAAAGCAT	3855
Db	1354	LysIleuLeuGluGIlyLysAsnGIyArGIThrValGIlyGIcLysAsnProLyGIUser	1373
OY	3856	AGTAATGGATTCGCCGANTGGCAATGC---GGAATATAAATTAGTGGAAGACCCTCGT	3912
Db	1374	-----TyPProAsPTPrAPScylsAsnSnIleAsPIIeSerHisASP---GIy	1389
OY	3913	GTGTGATGCCCCCTTAGAACGAAAAGTATGCGTACATTTCTTGGAAMATGATATGAA	3972
Db	1390	AlAcysMeTrProProArGIyArGIInLyLeucCysLeuTYrTLieAlaHisGLUserGIIn	1409
OY	3973	ATAAAAAATTACATCATCAAAGTTAAATTTAAAAAGACCTTTTCATCAANTCGCAGCAGA	4032
Db	1410	ThGIuSnIleIySThrAPSPASnLeuLysAspAlaPeIIeIySThrAlaIala	1429
OY	4033	GAAACATTCCTTCATGGTATTAATTATTAAGTAAG---GATGGTGAAGAAATGACATC	4089
Db	1430	GIuThrPhelSeuSerGIyInLYrTYrLysSerLysAsnSpserGIuAlaIySIIeU	1449
OY	4090	GATAAAGAATTAAGAAGCAAAATTCCTCCCGCATTTTGGATTCATGTTCTACACA	4149
Db	1450	AsprIG-----GIyleuIIeIleProSerGIInPheLeuarISerMetcetyThr	1465
OY	4150	TTTGGAGATTATAGAGATTTTTTATTTTGGAAAGATPATCAAAAGGTCAATGTGGAGGA	4209
Db	1466	PheGIyAsprLyArGIasprILecysLeuAsnThrAspIleSerLysGIuSnAsnAspVal	1485
OY	4210	AGTAACCTAAAGAAGCAAAATAGATTCCTTTTCAAAAATGGTGACCAAAAANTCCTAAT	4269
Db	1486	AlIsyAlaIalysAspLysIIeGIyLysPhePheSerLysAspLIySerLysSerProSer	1505
OY	4270	GGAAAAACAGCCAGAGATGGTGGACAGAAATAGCATGATGAGTATGGGAAGCTATGCA	4329
Db	1506	GIyLeuSerArGIInGIuTrPrLySThrAsnGIyProGIuIIeTrLysGIcMytCue	1525
OY	4330	TGTGCACTAGTAATAATTGGGGCAAAAAAGATGAT-----TTTACCGAAAACCTGAT	4380
Db	1526	CysAlaLeuThrLysTyGIyValThrAspThrAspSnLIyAsrGIySIIeLysAsnAsprYr	1545
OY	4381	GGTTACACAGACGTCAAATTTAGTGCAGAA--AGACACACTTTGGAGGAATTTGCCAAA	4437
Db	1546	SerTYrAsprLyAsnIleIeLysAspAlacysAsnGIuIIeAsnSerThrgInIncysAsn	1565
OY	4438	GCACCCCAATTTTTCAGATGCGCTAACCGAATGGTACGACGACTATTTCTATACAGACAA	4497
Db	1566	LysProGIInPheLeuArGIrPmetIIeGIyTrGIyGIInGIuPhecysAlaGIuArGIIn	1585
OY	4498	AAATATTTGAAGAGATCGACGAAAAATGTATGCTA---AATGACCAATGAAGTGTAT	4554
Db	1586	LysLysGIuAsnIIeIeLysAspAlacysAsnGIuIIeAsnSerThrgInIncysAsn	1605
OY	4555	ACA-----GAATGTATTAAGAAATCGAGACTACGCTTAATATTTGAAA---AAA	4602
Db	1606	AspAlaLysHisArGIcysAsnGIInlaCysArGIatIyArGIInGIuTYrValGIuSnIyS	1625
OY	4603	AAAAAAGGTGATTCSCCAATATTAATTAATTTCAAGAGATGAACGGCACAAAAAAGATTC	4662
Db	1626	LysLysGIuPheSerGIyGIInThrAsnAsnPreValIleuLysAlaAsnValGIInProGIIn	1645
OY	4663	GATACACAAACATTTGGTGTATGGTTCACAGACTACTACTGGAACGAATGCAACAGATTAC	4722
Db	1646	AsprGIuTYrLysGIyLysGIuTYrGIuTYrLysAsp-----	1656
OY	4723	TTGAACAGAAATTTAGTGTCTAGTGTGTGTATGAACCTGGAAGTGCCTGTGTGTACAA	4782
Db	1657	-----GIyAlaGIInPro-----IIeGIIn	1662
OY	4783	AGAAATATACATTTGTGA AAAACAGGCTACTACTATGTATGGCCGACAAATTTGGGTGC	4842
Db	1663	GIyAsnGIuTYrLeuLeuGIInLys-----CysAspAsnAsnLys---CysSerLys	1678

QY	4843	ACAAAATTATTTAGAAATGAGCACAATAATACTAACATTTGGAGTAAAGATTAAGTCAAA	4902
Db	1678	-----	1678
QY	4903	GGATTAGTAAAGGAGCAAAACACAGTGTCTATTAGTGGCAAAACAAAGTCTAATTAC	4962
Db	1678	-----	1678
QY	4963	TACATAACTTGAAGAATTGTACTGAGAGATGTCTTTTCTCTCTCGACTACGTATA	5022
Db	1678	-----	1678
QY	5023	TGTTTCATGCATTGATGGCAATTAATACAGATCCAGAGTTAAAGATGAATAATGGGTG	5082
Db	1679	----- ::: -MetaspGlyasnValIeuseryValSerProIysGluLys-----	1691
QY	5083	CGAAAAGATTGATGGAATGGCGGCAACGGMAGGTACATTTGGGTCAATACTACAAA	5142
Db	1692	----- -ProheGlyLysTy-----	1696
QY	5143	GAIAAAAAAGAAAAAGAAATATAAAACGTGCGATCGCACAAATTTCTTAGAGGTC	5202
Db	1697	----- -AlahIstysTyPro-----	1701
QY	5203	CCGCTTGTAGTGTCTATGAATAATAGTTTATGATTTAAGAGATTAAATTCAGTATT	5262
Db	1701	-----	1701
QY	5263	GATTAATTTGGAAGATGAAAAACAAAAGACCGAGAAAATTTGAAGAAAATATTTAACAA	5322
Db	1701	-----	1701
QY	5323	AATGGACATCAGTTGGCAAAAGNAGTAGTACTACAGAAATCCCGTAGTACTGCG	5382
Db	1701	-----	1701
QY	5383	CGAAATTTTCTTGGAACGAATAATAGGAATGTGTGTGGAACGCAATGATATCGGGTAC	5442
Db	1701	-----	1701
QY	5443	AAACGTGTAGGAGTATGGAATAATAGTGAATAAGTCAAGAGTATGAAGATCTMAAA	5502
Db	1701	-----	1701
QY	5503	AAATGTGTTCTTACTCTCAGATGATATTATCTATGGGAAAAATCGATGAAGGT	5562
Db	1701	-----	1701
QY	5563	ACTGCTATCAGTTTCTTCGATGTTTGGCGAATGGGGGTGAAGATTTTTCGAACATAAA	5622
Db	1701	-----	1701
QY	5623	GAIAAGAAITGGAGAATTTGTAGGGGCGGTAAATGATTATACTTGTGTATATGAA	5682
Db	1701	-----	1701
QY	5683	GATAAAGAAGAAATGTACAGATGCGTGTACACAAATATAAAAAATTTATAGTAGTGG	5742
Db	1702	----- ::: -GluLysCys-----	1704
QY	5743	AAACACAGTATGAIAAACCAAAATCAAAAAATATGTGAGAAATAAAGACAAATATATCC	5802
Db	1704	-----	1704
QY	5803	GAGCATCTGTGGCAAAAGATCAGAGACGCTGCGGAATATTTAGACAAACAATTAATA	5862
Db	1704	-----	1704
QY	5863	AAAAATTTGGAATAAAGTGGAGATTGTGAATATAAGTATGAAGAAGTGTGCCACA	5922
Db	1705	----- -AspCys-----	1706
QY	5923	CAGCGATTACTGATGATTAATGTCAAAATATGCCCGCATCATTTAGACATGAACAAA	5982

Db 1706 ----- 1706
QY 5983 GAAGTTGAAGAAAGTAAATGTCAGAGTCACAGAGTCACAGAGTACAGAGGAA 6042
Db 1707 ---TyrGlnGlyLys-----HisValProSerIleProPro----- 1717
QY 6043 ACACCGTCACACGAGGTATCTGATATCAAAAGCGAGCATCAAAAGAGCGAAA 6102
Db 1717 ----- 1717
QY 6103 ACAGCGCCCTCAACAAAGCCGAAAGTGAATAATCAACACAGAAATGCCAGCA 6162
Db 1718 ---ProProProProValGlnProGln----- 1725
QY 6163 CAACACGAAACCCGACGAGCAGCAACAAACAGAAACGACATCAACAGCAGCAACCA 6222
Db 1726 ----- 1730
QY 6223 ACAGAAATCTGACGTGGGACCAATGTAAAGCCATCTTCGAATAAACACAGATACAG 6282
Db 1731 ValThrValAspValLysSerIleValLysThrLeuPheLys-----AspThrAsn 1747
QY 6283 GGTGGAATAGAGGTTGTATCCAAACGATGACCAATAT-----CCTAAATGGGGT 6336
Db 1748 AsnPheSerAspAlaCysGlyLeuLys---TyrGlyLysThrAlaProSerSerThrLys 1766
QY 6337 TGTATT-----GTAGTAAAGTCTAAAGAAAT 6363
Db 1767 CysIleProSerAspThrLysSerGlyAlaGlyAlaThrThrLysSerGlySerAsp 1786
QY 6364 GAAATATGCAATATGATGCTCTACAGAGAAATAATATGATTAATAATATATACATAT 6423
Db 1787 SerGlySerIleCysIleProProArgArgArgLeuTyrValGlyLysLeuGln--- 1805
QY 6424 TTAATATATGCAACT-----GAAATATAGCGTGAC 6453
Db 1806 ---GlnThrAlaThrAlaLeuProGlnGlyGlyAlaAlaProSerHisSerArgAla 1824
QY 6454 AATGATATATAAGAGCTTTTATTAATGTCAGCAATAGCAATCAATTTTGTGTAT 6513
Db 1825 AspAspLeuArgAsnAlaPheIleGlnSerAlaAlaIleGlnThrPhePheLeuThrPasp 1844
QY 6514 AAATATATATATGAA---AATCTGACAGCAAAATGAA----- 6549
Db 1845 ArgTyrLysGlnGlyLysLysProGlnGlyAspGlySerGlnGlnAlaLeuSerGlnLeu 1864
QY 6550 -----TTGCAAAATGCAACA 6564
Db 1865 ThrSerThrTyrSerAspAspGlnGlyAspProProAspLysLeuGlnAsnGlyLys 1884
QY 6565 ATTCCAGATGAAATTTAAAGAAATATGATATATATATGATGATTTTAAAGATGTTT 6624
Db 1885 IleProProAspPheLeuAlaGlyLeuMetPheTyrThrLeuGlyAspTyrArgAspIleLeu 1904
QY 6625 TTTTGAACATGATTTTCTTAATGATTAATAAAATATATTAACATCAAAATAGCTGTACACCC 6684
Db 1905 ValHisGlyGlyAsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnIle 1924
QY 6685 ATTCTCAATGAAATATATAGAAAGAAACGACATTAATAAAATATCAAGATTAACGAAA 6744
Db 1925 ValLeuGlnAlaSerGlyLysLysGlnAspMetGlnLysIleGlnLysIleGlnGln 1944
QY 6745 ATA----- 6747
Db 1945 IleLeuProLysAsnGlyLysThrProLeuValProLysSerSerAlaGlnThrProAsp 1964
QY 6748 ---TTTGGAGCAAAATATAAAATTTATTTGGAGAGCAATATATGATGATTAACCTTAT 6804
Db 1965 LysTrpTrpAsnGlnHisAlaGlnSerIleTrpLysGlyMetIleCysAlaLeuThrTyr 1984
QY 6805 -----CATCTCACAGCAAGCAAAACGAAAGAAATTTAGA 6840
Db 1985 ThrGlnLysAsnProAspThrSerAlaArgGlyAspGlnLysLysIleGlnLysAspAsp 2004
QY 6841 GATTAATAC----- 6849
Db 2005 GlnValTyrGlnLysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrPro 2024
QY 6850 -----CAGTACAAAT---GACATGACCAAACTG----- 6873
Db 2025 ThrGlyThrTyrLysThrGlnTyrAspTyrGlnLysValLysLeuGlnLysPheThrSerGly 2044
QY 6874 -----ACGCTTCCTCCCTTGAAAGCTTTGAAAAAGCCC 6906
Db 2045 AlaLysThrProSerAlaSerSerAspThrProLeuLeuSerAspPheValLeuAspPro 2064
QY 6907 CAATTTTGAGATGTTCAAGAAATGGCAGAAATTTTGTATTAATGAAGAGAAAGCAAG 6966
Db 2065 ProTyrPheArgTyrLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2084
QY 6967 TTTGTTAAATTTGAGAGCGGCGCTGTAAAGAAATATGAG----- 7002
Db 2085 LeuAlaGlnIleLysHisGlnCysLysValGlnGlnLysAsnGlyLysGlySerArgArgGly 2104
QY 7003 -----TGTAAAT-----GGTAAAT 7017
Db 2105 GlyIleThrArgGlnTyrSerGlyAspGlyGlyAlaCysAsnGlnLeuProLysAsn 2124
QY 7018 GACGTAAGACACAGAA-----TGTGACAGAGCGGTGTATACATATCA 7062
Db 2125 AspGlyThrValProAspLeuGlnLysProSerCysAlaLysProCysSerSerTyrArg 2144
QY 7063 AATTTTATTAAGAGTGAAGAAACTGAATATGAAGAAACAGCAAAAGTTCAAAAGAT 7122
Db 2145 LysTrpIleGlnSerLysGlyGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 2164
QY 7123 AAAGAT----- 7128
Db 2165 LysAspLysCysValAsnGlySerAsnLysHisAspAsnGlyPheCysGlnThrLeuThr 2184
QY 7129 ---GGCAAAAGATATAGATATAT-----CCTTCTACTGAAGAGACATA 7170
Db 2185 ThrSerSerLysAlaLysAspPheLeuLysThrLeuGlnProCysLysProAsnAsnVal 2204
QY 7171 GACAGGCAACATGCTCTCATGATATTAACATGAATAATTAAGAAATATGTCGCAAT 7230
Db 2205 GlnGlyLysThrIlePheAspAspLysThrPheLys-----HisThr 2219
QY 7231 AAGGATTTGTTCTGATGCAAAACCTTCTTCAACATACCAAAACACACAACATCA 7290
Db 2220 LysAspCysAspProCysLeuLysPheSerValAsnCysLysLys----- 2234
QY 7291 CAATCATCCGATGCTATGATATGACCAATCGCTGATATATGTCGTAAGAAATTAAC 7350
Db 2235 -----Asp 2235
QY 7351 AAGTGTAGTCTCTCAACTTTCAAAAGAGATCTATGATCATACAAAAAATTAATCT 7410
Db 2236 GlnCysAsp-----AsnSerLysGlyThrAspCysArgAsnLys----- 2248
QY 7411 GAACCTAAATATACCTATGAAATTTGTGAAGAAAGACATATATATCTATAAGAACGA 7470
Db 2249 -----AsnSerIleAsp-----AlaThrPheIle 2256
QY 7471 GAAATATATGATATTAATCTTGAAGAGAAATTTATATACCTATGATGATCTACAAAGAA 7530
Db 2257 GlnAsnGlyValAspSerThrVal-----LeuGlnMetArgValSer 2270
QY 7531 AACGAAGTAAATAATAGTGGACTAATATATATCTTGGATCTTAAGAAACCTTATGCA 7590
Db 2271 AlaAspSerLysSerGlyPheAsnGlyAsp----- 2280
QY 7591 CCGATTAATATATATAGGAAGCAAGAACCCCTTGAAAT----- 7629
Db 2281 -----GlyLeuGlnAsnAlaCysArgGlyAlaGlyIlePheGlnGlyIle 2295

[illegible]

QY	4102	AAAGAGCAAAATTCCTCCGATTTTGGATTCATCCATGCTTCTACACATTTGGAGATTAT	4101
Db	1445	-----GlyLeuIIleProSerGlnPheLeuIaValSerMetCtyTrpPheIlyAspIyr	1462
QY	4162	AGAGATTTTATTTATTTGGACAGATATATATCAAAAGCATGCTAGAGCAAGTAACATAAA	4221
Db	1463	ArgAspIleCysLeuAsnThrAspIleSerLysLysGlnAsnAspValAlaLysAlaLys	1482
QY	4222	GAGCAATAGATTCTCTTTCCAAAATGGTGACCAAAAATCTCTTAATGGAAAAACAGC	4281
Db	1483	AspLysIIleGlyLysPhePheSerLysAspGlySerLysSerProSerClyLeuSerArg	1502
QY	4282	CAAGAAATGGCGACAGAAACATCTCATGTAGATATGGAGCAATGCTATNGCATACTA	4341
Db	1503	GlnGluTrpIlePyrLysThrAsnGlyProGlnIleTrpLysGlyMetLeuCysAlaLeuThr	1522
QY	4342	AAATTTGGCGCAAAAAAAGATGAT-----TTTACGCAAAACCTACGGTTTACACAC	4392
Db	1523	LysTrpValThrAspThrAspAsnLysArgLysIIleLysAsnAspLysSerTrpAspLys	1542
QY	4393	GTCAAATTAGTGACAAA---ACACCACTTTGGAGAAATTTGCCAAACACCCCACTTT	4449
Db	1543	ValAsnGlnSerGlnAsnGlyAsnProSerLeuGlnGluPheAlaLysProGlnPhe	1562
QY	4450	TTACAGGCGCTAACCCGATAGTGTACGACGACTATGCTATACACGCAAAAAATTTTGA	4509
Db	1563	LeuArgTrpMetIIleGluTrpGlyGlnGluPheCysAlaGlnArgGlnLysGluAsn	1582
QY	4510	GATGTGCAGCAAAAAATGTAAATCA---AATGACCAATTTGAAGTGATACAC-----	4557
Db	1583	IIleLysAspAlaCysAsnGluIleAsnSerThrGlnGlnCysAsnAspAlaLysHis	1602
QY	4558	GATGTATATAGAAATGCGAGCACTCGTTAAATATATGAAA---AAAAAAAAGATGG	4614
Db	1603	ArgGlyAsnGlnAlaCysAlaGalaTrpGlnGluTrpValGlnAsnLysLysGluPhe	1622
QY	4615	ATTCACACAGATTAATATTTACAAGATGAACCGCAAAAAAAGATTCATAGACACAC	4674
Db	1623	SerGlyGlnThrAsnAsnPheValIIleLysAlaAsnValGlnProGlnAspProGluTr	1642
QY	4675	ATTTGGTGAATCGTTACAGACTTACTGGACACAAATGCAACAGATTACTTGACACGAAA	4734
Db	1643	LysGlyLyrGluTrpLysAsp-----	1649
QY	4735	TTTACTGCTAGTTGGTGTGTAATCCCTGGAAAGTGCCTGTGTACAAAGAAATATACA	4794
Db	1650	-----GlyValGlnPro-----IIleGlnGlyAsnGluTrp	1659
QY	4795	TTGTTCAGAAAAACAGCGCTTCTATCTATGCCGCAACAACATTGGGGTGCACAAATTTAT	4854
Db	1660	LeuLeuGlnLys-----CysAspAsnAsnLys---CysSerCys-----	1671
QY	4855	GAATATGACGCAAAATATCTAATCACTTTGCAGTAAAGATTAAGTCGAAAGATTAGTAAG	4914
Db	1671	-----	1671
QY	4915	GAGCAAAACACAGTGCTATTAAATGGCAAAAACAAAGTCTTAATATCAATAACTTG	4974
Db	1671	-----	1671
QY	4975	AAAGATTTGACTGAAGATGTGCTTTTCCCTTCGTCGACTACGTATATGTTTTCATGCA	5034
Db	1671	-----	1671
QY	5035	TTTGATGGCAATTATACAGATCCAGAAATTAAAGATGAATAATGGGTGGCAAAAAAGATTG	5094
Db	1672	MetAspGlyAsnValIleuSerValSerProLysGluLys-----	1684
QY	5095	ATGGAATGGCGCGCAACGGAAGGTACAAATTGGTGTCATCTACTACAAAAGAAAAAGAA	5154
Db	1685	-----ProPheGlyLysTrp-----	1689

QY	5135	AAAGGAAAAATAAAAACGTGGCATCGGCCCAAAATATCTTATAGAGTCCCGCTGTAGT	5214
Db	1690	-----AlaHisLysTyrPro-----	1694
QY	5215	GCTATGAATATAGTCTTTTATGATTTTAAGATATATCTAGTATTGATTAATTGGAA	5274
Db	1694	-----	1694
QY	5275	GATGAAAAACAAAAGACCAGAAAAATTGGAMGMAATATTTAACAAAAATGGAAACATCA	5334
Db	1694	-----	1694
QY	5335	GTTGGCAAAAGAACTGATAGTACTACAGAAATCCGGTAGTACTCGCGAAAAATTTTTC	5394
Db	1694	-----	1694
QY	5395	TGGAAAGAAATTAAGAAATGTGTGTGGAAACGCAATATATGGGGGTACAAAAGTGTAGG	5454
Db	1694	-----	1694
QY	5455	GATGTGGAATAGTGGAAATAGTGCMAAGATGATGAATCTAAAAAATGTGTTCT	5514
Db	1694	-----	1694
QY	5515	GTACCTTCAGATGATGATATTCCTATGGGAAAAATCGGATGAAGTACTGCGTATCAG	5574
Db	1694	-----	1694
QY	5575	TTTCTGATGTGTTGGCCGAATGGGGTGAAGATTTTTCMAAACATTAAGAAAAAGAAATTG	5634
Db	1694	-----	1694
QY	5635	GAGAAATTGTAGGGGCGGTGAATGATTTACTTGTGTATATGAAGATAAAGAAAG	5694
Db	1695	-----Glu	1695
QY	5695	AAATGTACAGTCGGCTGTACAAATATAAAAAATTTATAGTAGAAGAACACAGTAT	5754
Db	1696	-----LysCys	1697
QY	5755	GAAAAACAATCAAAAAATATGTGGAATAAAGACAAAATATATTCGAGCATCTGTG	5814
Db	1697	-----	1697
QY	5815	GCAAAAGATCGAGAGACGCTCGGAATATTTAGACAAACAATTAAAAAATTTGTGA	5874
Db	1697	-----	1697
QY	5875	AATAAAAGTGAGATTGTGAATATAAGTATGATAAAGATGTGTCCACACGCGATTAACT	5934
Db	1698	-----AspCys	1699
QY	5935	GATGTAATAGTCAAAAATATGCGCGCATCTTAGACAGATGAACCAAAAGAACTTACGA	5994
Db	1700	-----TyrGlnGly	1702
QY	5995	AAGTAAATTGTCAAGTCCGACGAGGTCCACACGCTGTACGAAGGGAAACACGTCACCA	6054
Db	1703	-----HisValProSerIleProPro	1710
QY	6055	CGGGATCACTGATATCAAAAAGCGACGCGATCGAAAAAAGAACGCAACACCGCCGCT	6114
Db	1711	-----ProProPro	1713
QY	6115	ACAAAACGCCGAAAAAATGTGAAATCTTAACAAGAAATGCGAGCAACAACAGCAACC	6174
Db	1714	-----ProValGlnProGln	1718
QY	6175	CGACGACGACGACAACAACACGAAAAACGAATCAACACGACAACAACAGATCTGAC	6234
Db	1719	-----ProGlnAlaIleProThrValThrValAsp	1727
QY	6235	GTGGGCAATAGTAAAGCCATTCTTTCGAATAAACAGATAGCAGGGGTGGAATAGAG	6294

Db	1728	ValCysSerIleValLysThrLeuPheLys-----AspThrAsnAsnPheSerAsp	1744
QY	6295	GGTGTGAATCCAAAAACGTAATGACAAATAT-----CTTAATAGGGGGTGTATT-----	6342
Db	1745	AlaCysGlyLeuLys---TyrGlyLysThrAlaProSerSerThrPheCysIleProSer	1763
QY	6343	-----GTAGCAATAGCTTAAGAAATATGCAAAATGCAATATGCAAT	6375
Db	1764	AspThrLysSerGlyAlaGlyAlaThrThrGlyLysSerGlySerAspSerGlySerIle	1793
QY	6376	TGTATGCGCTCCTAGAGAAAAAAATATGTAATAAATATATACAAATATTTAAATATGAA	6435
Db	1784	CysIleProProArgArgArgArgLeuTyrValGlyLysLeuGln-----GluThrPheAla	1801
QY	6436	ACT-----GAAATATACCGTGCACAAATGATATATAA	6465
Db	1802	ThrAlaLeuProGlnGlnGlyGlnGlyAlaAlaProSerHisSerArgAlaAspSerLeuArg	1821
QY	6466	GAGCGCTTTATTTAAATGTGCACAAATAGAAATCAATTTTGTGCTTAAATATTAAT	6525
Db	1822	AsnAlaPheIleGlnSerAlaAlaIleGluThrPhePheLeuThrPheAspArgTyrIleSerL	1841
QY	6526	GAA-----AATCCTGCACGACGAAAAATGAA-----	6549
Db	1842	GluLysLysProGlnGlyAspGlySerGlnGlnAlaLeuSerGlnLeuThrSerThrTyr	1861
QY	6550	-----TTGCAAATATGTAAGCAATTCGACGATGAA	6576
Db	1862	SerAspAspGluGluAspProProAspLysLeuLeuGlnHisGlyLysIleProProasp	1881
QY	6577	TTTAAACAAATATATGATTTATACATATGGTGAATTAAGATATATGTTTTTGGAACTGAT	6636
Db	1882	PheLeuAspGluMetPheTyrThrLeuGlyAspTyrArgAspIleLeuValHisGlyGly	1901
QY	6637	ATTTCCTATGATAAAAAATTTATACGTGAACAAATAGCTGAACAACATCTCGATGAA	6696
Db	1902	AsnThrSerAspSerGlyAsnThrAsnGlySerAsnAsnAsnAsnIleValLeuGluAla	1921
QY	6697	AATATATGAAGAAAAACAGGATTAAGAAAAAAGATGAAGATTAACGTAAATA-----	6747
Db	1922	SerGlyAsnLysGluAspMetGlnLysIleGlnGlnLysIleGlnIleLeuProLys	1941
QY	6748	-----TTTTGGGAC	6756
Db	1942	AsnGlyGlyThrProLeuValProLysSerSerAlaGlnThrProAspLysTrpTrpAsn	1961
QY	6757	AAAAATAAAAATTTATTTGGAGGAAGATGATATGATTAACGTTAT-----	6804
Db	1962	GluHisAlaGlnSerIleTrpLysGlyMetIleCysAlaLeuThrTyrThrGluLysAsn	1981
QY	6805	-----CATCCACAGACGAAAAACGAAAAAGAAAAAATTAGAGATATTAC---	6849
Db	1982	ProAspThrSerAlaArgGlyAspGlnAsnLysIleGluLysAspAspGluValTyrGlu	2001
QY	6849	-----	6849
Db	2002	LysPhePheGlySerThrAlaAspLysHisGlyThrAlaSerThrProThrGlyThrTyr	2021
QY	6850	-----CAGTACAT---GACATGACCAAACTG-----	6873
Db	2022	LysThrGlnTyrAspLysPheGlnLysValLysLeuGlnAspThrSerGlyAlaLysThrPro	2041
QY	6874	-----ACGGCTTCCTGAGAAAGTTTGTAATAAAGCCCAATTTTGAGA	6918
Db	2042	SerLysSerSerAspThrProLeuLeuSerSerPheValLeuArgProTyrPheArg	2061
QY	6919	TGGTTCCAGCAATGGCAGACAAATTTTGTATAAGAGAGAGAACAGTGTAAATATG	6978
Db	2062	TyrLeuGlnGluTyrPheGlyLysAsnPheCysLysLysArgLysHisLysLeuAlaGlnIle	2081
QY	6979	GAGCGCGGCTGTAGGAATATGAG-----	7002

Db	2082	LYNHISGLUCYLSELYVALGICLUASNDLYLYLYSETARGGLYLIIEHTRG	2107
QY	7003	-----TGTAAT-----GGTAGTAAGCGGTAAAGCA	7029
Db	2102	GLNTYSErgLYASpGLYGLIALACYSANDLIUETLEUPROLYSASNDPGLYTRVAL	2125
QY	7030	CAAGAA-----TGACAGAGCGGTGTATCAATACAAATTTATTAAG	7074
Db	2122	PROASPLEUGLILYRPROSECYALALYSPROCYSESErTYARGLYTRILEGU	2149
QY	7075	AACTGAAAACCTGAATATGAAAGCAAAAGAAAAGTTCAAAAAGGATTAAGT	7128
Db	2142	SErLYSGLYLUPHEGLILUGLIGLILUVALATYGLICLUINDLILMSpLYSCYS	2161
QY	7129	-----GGCAAAAG	7137
Db	2162	VALASNGLYSErASNDLYSHISASpASNGLYPHECYSGLILHrLEUTHrHrSErSErLYS	2181
QY	7138	TATPAAGATTAAT-----CCCTTCTACTGAAAAGACACATACAGAAAGGCACCA	7182
Db	2182	ALALYASpPHELEULYETHrLEUGLYPRCCYLYSPROASMSVALGICLUILYSTR	2201
QY	7183	TGTGCTCATGATATTTAAACATGAAATTTAAAGAAATTATGTGGCAATAGAGTTGTTCT	7242
Db	2202	LLEPHEASpASpLYSTRHrPHELYS-----HISrHYLYSAPCYSPASp	2216
QY	7243	TGTATGCAAAAACCTCTTCCACACTCCACAAAACACACACACATCATCATCCGAT	7302
Db	2217	PROCYSELYLYSPHESErVALASCYSLYLYS-----	2227
QY	7303	GCTAATGATATGCCAGAAATCGTGGAATATGTTCTCGAAGAAATTTAAACAGTGTGAGTGT	7362
Db	2228	-----ASpGLUCYASp-----	2231
QY	7363	CCTGAACCTTCAAAAAGGCTCATGATCATACAAAAAATTACGAACCTAAATA	7422
Db	2232	-----ASerLYSGLYTHrASpCYSArGASNDLYS-----	2241
QY	7423	CCTATGAATGTGTAGAGAAAAGCAGCATATTTATTTAAAGAACGAAATATATATG	7482
Db	2242	-----ASerLIEAsp-----ALATHrASpLIEGLUASNDLYVAL	2253
QY	7483	GATATTACCTTGAAAGAAAATTATTAACCATTTGAGTCTACAAAGAAAAGCAAGTAA	7542
Db	2254	ASpSERHrVAL-----LEUGLMErVALSERVALASpSERLYS	2267
QY	7543	AATAGTTGGACTAATATATTAATCTTCGACATCTTAAGAAAACCTTATGCACGTGATTAATAT	7602
Db	2268	SErGLYPHEASNDLYSP-----	2273
QY	7603	ATAGAAAGAAAACCTGTGTGAAT-----AGAGAAAGAAAT	7641
Db	2274	---GLYLEUGLUBENALACYSArGGLYALAGLYLIEPHEGLUGLYLIErGLYASp---	2291
QY	7642	CGTTTTAAAGTAGATTTATGTAAGTAAGTAATTTCAAAATTTCAAAAGTTCCTACAGAGAAA	7701
Db	2292	-----GLUTRILYSCYS---ARASNDVALCYSGLYTYrVAL-----	2302
QY	7702	AAAAAGTATGTGACTCTCCACAGACAGACATATGTCTTAAGAAATTTAGATGAATT	7761
Db	2303	-----VALCYSLYSPRO-----	2306
QY	7762	AAATGTAAAGACTTAAAGALNGTATATATCTCCATAAAATGGTTGTCGAACTGCACGA	7821
Db	2306	-----	2306
QY	7822	AATGAAGAAATAGACATATTAATAAAACCTTCAACTCAGAGAACGGGTGCGCAATGAATCA	7881
Db	2307	-----GLUBENVALASNDLYGLU-----	2312
QY	7882	ATATGTATACTGAATAATATAGTTTCCTGTATCTGGGTGACATAGTTTAAGACAAACAT	7941
Db	2313	-----ALALYSGLYLYSHIS	2317


```

Db      84  LeuLysGlyIleThrAsnGlnGluArgPheSerAspThrLeuGlyGlnCysThrAsn 103
QY      382  AAAATA-----CGAATTTATAAAAGAAAAAT-----GATGGTATAGCTGCGC 426
Db      104  LysLysIleLysGlnGluArgPheSerThrLeuSerGlyLysAspCysGlyAlaCysAla 123
QY      427  CCACCTAGAAAGACGACATATGTGTGATATAAACTTGAGACCTGTAATGATATAATACC 486
Db      124  ProTyrArgArgLeuHisLeuCysSerHisAsnLeuGlnSerIleAsp-----ThrThr 141
QY      487  CAAAATTCATGATTTATGTTGGAAAAAGTACTAGTACACCAAAATTCGAAGTGAATCA 546
Db      142  SerMetThrHisLysLeuLeuLeuGlnValCysMetAlaIleLysTyrGlnLysAsnSer 161
QY      547  ATTGTTAATATGATCA-----CATAAAGAACT-----TCAGACGGCTGT 588
Db      162  IleAspThrHisTyrProGlnHisGlnArgThrAsnGlnAspSerProSerGlnIleCys 181
QY      589  ACTGCTCTTCACGAACTTTTGCAGATATAGTGATTTGTAAGAAATAGATATGTT 648
Db      182  ThrMetLeuIleAspSerPheAlaAspIleGlyAspIleValArgGlyLysAspLeuPhe 201
QY      649  AAACCAAT-----GTCCATGACAAAGTAAACGGGCTCCGAGAGCTT 693
Db      202  TyrGlyAsnSerLysGlnLysGlyLysArgAspGlnLeuGlnThrAsnLeuLysThrIle 221
QY      694  TTCAGAAATATACATGATGATGAATGAAGATGAA-----GTAAATAATGATTAATCCCT 747
Db      222  PheGlyLysIleHisGlnLysLeuLysAspLysGlnGlyAlaGlnThrArgTyrGlySer 241
QY      748  GATGATCTCGGAAATTTATTAATTAAGAAACATGCTGTGAATGTGAATAGAAATAA 807
Db      242  Asp---ThrThrasnTyrGlnLeuArgGlnAspTyrPyrAlaAsnArgAlaThr 260
QY      808  GRTGGGAAGCTATACATGATGATGATCATATTAATCGATTTATTTATGCAATCA--- 864
Db      261  ValTyrGlnAlaIleThrCysAspVal---HisGlySerAspTyrPheArgGlnThrCys 279
QY      865  -----GAAAGTAAATACACCATTAATTTCAATCTTAATGTC----- 900
Db      280  GlyAspLysGlnThrThrAlaThrArgValLysAspLysCysArgCysLysAspGlnAsn 299
QY      901  GGGCATAAACAAGCA-----AAGTCTCCACCAATTTAGATATATGCTCCCTCA 948
Db      300  GlyLysLysProGlySerAsnAlaAspGlnValProThrTyrPheAspTyrAlaProGln 319
QY      949  TATTTACGTTGGTTCGACGATGGGGAAGAGTTTGGCGAAAAATTTAAATG 1008
Db      320  TyrLeuArgTyrPheGlnGlnIleThrAlaGlnAspPheCysArgLysLysLysLeu 339
QY      1009  AAAAAGTCAAGACCTCTGCTGTAATGACAAAGAACGCTTATATTGATGATAGCA 1068
Db      340  GlnLysLeuGlnGlnGlnCysArgAspTyrLysGlnAsnLeuTyrCysSerGlyAsnGly 359
QY      1069  CATGATTTGACGACAACTATTGGAAGAAAGTATTGCACTTTGATGATTAAGTACT 1128
Db      360  TyrAspCysThrLysThrIleTyrLysGlyLysLeuValIleGlyClnHisCysThr 379
QY      1129  GACTGTTCGACTAAATGCAAAAGTTTGGTAAGTTGTTAGGAAATCAACAAGACATTT 1188
Db      380  AsnGlySerValTyrCysArgLeuTyrGlnSerTrpIleAspAsnGlnLysLeuIlePhe 399
QY      1189  AAAAAACAAGAAATATGAAAAAGAAATACATATATTATTCG----- 1236
Db      400  LeuLysGlnLysGlnLysTyrGlnThrGlnIleSerAsnSerGlySerCysGlyLysSer 419
QY      1237  -----ACGATAAACAATTTGTCATATATATTAATAGTAA----- 1272
Db      420  GlyGlyValLysGlyArgAsnArgLysArgGlyAlaGlyAlaGlnThrAlaThrAsn 439
QY      1273  -----TATTAATAACAATTTTGAATAAACTTAAGAAACGCAATATGCAACTAAT 1323

```

```

Db      440  TyrAspGlyTyrGlnLysLysPheTyrLysGlnLeuLysGlnSerGlnTyrGlyLysVal 459
QY      1324  GACACTTTTAAATTTACTAAATGAAGAAAGATATTGTAAGAGAGATATACAGACAA 1383
Db      460  AspAspPheLeuLysLeuLeuAsnGlnAlaAspValCysLysLysIleLysAspGlnLys 479
QY      1384  AAGGATATTAATCTTTACTTAACAGTGTGAT-----GACAAAGGATATTATT 1431
Db      480  GlnLysIleAspPheThrLysProAlaAspLysAsnSerAsnAsnGlnIleThrPheTyr 499
QY      1432  CGTTCAAGATTAATTCGCAAGCTGTCCCGACCTCGGGCTCAAAATGATGATTAATAATC 1491
Db      500  HisSerGlnTyrCysLysProCysProAspCysGlyValLysArgLysAspAsnGlnTrp 519
QY      1492  ACACAAATATCAGATATGATGCTGACGCTGAATTAATGAAGACTATTAACCTCCATGC 1551
Db      520  LysAspLysTyrAspGlyLysCysThrArg-----GlyLysLeuTyrGlnProAlaSer 537
QY      1552  GGTGTGAAGCTTACTAAATTTACATGCTCTTATAGTGTAATGAACAAGTATTTACA 1611
Db      538  GlyAlaGlnGlyThrProIleLysIleLeuLysSerGlyGlnLysGlnLysGlnIleGlu 557
QY      1612  CAAAAATTAAGAAATTTTGTAC-----ACCTCAACTAAT----- 1647
Db      558  ThrLysLeuLysAlaPheCysAspGlnThrAsnGlnLysAspThrThrAsnSerValAlaArg 577
QY      1648  -----TACAACATATAAAATATCAAAA-----TGG 1674
Db      578  GlyGlyGlyAlaAspGlySerGlySerLysSerAsnSerLysGlnLeuTyrGlnGlnTrp 597
QY      1675  GAATGCTATTATTAAG----- 1689
Db      598  LysCysTyrAsnGlnValGlnLysValLysAspAspLysAsnGlyGlnGlnAspGln 617
QY      1690  ---GATGAATATATAATAGA-----TGTAACTCGAA----- 1719
Db      618  AspGlnGlnAspValAspLysValLysAlaGlyLysLeuCysIleLeuGlnAsnLys 637
QY      1720  -----CAAAATACGAAATCAATTAATGATAATCTCAAGATATATCTTTGAT 1767
Db      638  LysHisGlnSerArgAsnAsnSerSerAsnGlnProGlnGlnPheGlnLysThrPheHis 657
QY      1768  AATTTTGTGATTAATGAGGTATACATATTATTAAGGACTACTATTAACTGAATGACAA 1827
Db      658  AspPheTyrPheTyrPheTrpIleGlyArgPheLeuAsnAspSerMetCylTrpArgGlyLys 677
QY      1828  CTTAATACTTGATATAATAAT---ACAACACGCAATTTGATGTAATGTAACGAAT 1884
Db      678  ValAsnSerCysIleAsnAsnProLysArgLysLysCysArgAsnGlnCysLysAspAsp 697
QY      1885  TCGTTATGTTTTCACAGATGGGTTAAACAAGAAAGAAAGATGCAATGATATTAAGAA 1944
Db      698  CysGlyCysPheLysGlnTrpIleGlyLysLysLysGlnGlnTrpGlnAsnIleLysLys 717
QY      1945  CTGTTCACAAAAAAAGAAATATACAGCAATCGTATTATGATTAATTAATCTTTT 2004
Db      718  HisPhe----- 719
QY      2005  GAAGGTTATTTTAAAGTTATGATTAACCTTACAAAGATGAGCAAAATGAAGAA 2064
Db      720  ----- 723
QY      2065  CTTATGAAATATATAAAGAAAAAATGATGTTTCAATTTGCAAAATTAATAGGAC 2124
Db      724  AlaPheLysAsn---LysArgGlnAsnSerGlyIleAspMetPheSerGlyLeuMetAsp 742
QY      2125  TATTTAAGAAATGCAATAGAACTCTTGTAGATCACTTAAGAAGAACTGCCAGATATGT 2184
Db      743  SerAlaAspValValLeuGlnLeuValLeu----- 752
QY      2185  AAGACATAATATCAAAAGCAAGCATGTGAAGATCCATTAATGACACAAACCCGTGT 2244
Db      752  ----- 752

```

```
QY 2245 GTTAAACCTCGTGAGGAGCGACCAACCTAATAATATAAGAAATAGACATACCTT 2304
Db 753 -----
QY 2305 AAAAGAGTGCATACGAGAGACGACGAATCGTGTCTCTCATTAATTGAAGAAAGCA 2364
Db 758 -----
QY 2365 CACGAGGTATATATAACGTGGGGTAGGAGAAAGACTCAAGACAAATTTATGTAGA 2424
Db 759 -----
QY 2425 ATAAATGATAAACATCTTAATCGTAATCTGTTTTTCAATGAGCACCATGTGATGCCAA 2484
Db 763 -----
QY 2485 GGCACAGTGTGTATACAAACAAGATTGTCTAGGAAGTGAAGTGGAAAGTGCCTG 2544
Db 764 GATTTAGTAAATATCTCTTTTGGGGAGTGTCTCTATACGCAAAATATGAAGCAAC 2724
QY 2725 AAGATATACGATGTATAAGAAAGATTAACCTAAGGGCCCAAGAAAGTAACTGAC 2784
Db 794 -----
QY 2795 -----
QY 2785 CCAAAACACGACACTATCTGTGACGTATACGTTTTCAGATATAGGTGAT 2844
Db 798 AsnGlnLysThrThrIleAspLysLeuLeuGlnHisGlu-----
QY 2845 ATATTTGAGAGAAAGATCTCTGTGGAAAGAAAGCGTGCATGTAAAGCTGCAGACAT 2904
Db 811 -----
QY 2905 TTGAAACGTGTTTGGTAAATATACATAGTCACTCAAGGCAAGAAATGATAAATAT 2964
Db 811 -----
QY 2965 AATGATGATGCCCAAAATATTTAAATTTGAGGAAAAATGTGGGAACTAATAGAGCC 3024
Db 812 GlysAspAlaAsnAsnCysLeuLysThrHis-----
QY 3025 AAAATATGGAAGCAATGATGATATATAAATTTGAAGATTAATCGGACACCAA 3084
Db 822 -----
QY 3085 TCAACACAAGTAGTATTGCGGATATAGTATCATACACATTGATGATTATATCCA 3144
Db 822 -----
QY 3145 CAAAATTAAGATGATGATGCCGAATGGCGAGATGTACTCAAGGTGCAGAAAAAGAG 3204
Db 822 -----
QY 3205 TATGATAGTTGAAGAGAAAGTGAAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 3265
Db 823 -----
QY 3266 CAAGGCTGTACGAAGAGAGTGTACAGGTTGTACGAAGTGCACAGAAAGTTGTAAAGA 3315
Db 839 ProGlyAlaProSerGlnThrGlyGlnThrThr-----
QY 3316 TATTAATGATTAATAGCATTAATGGAAGAAACAATGCAATTAATATACATAACAA 3375
Db 851 -----
QY 3376 GAATTCATGACAGACAAATGTCTGTATGTAATAGTGTGATGAACTTGCAGACT 3435
Db 857 GlnGlnAspGlnGlu-----
QY 3436 GCCAAAATATCATATAGACAGAAATGTATTAATTTTGTGGAAATTAATACCAAAAT 3495
Db 868 -----
QY 3496 GGTGGCAAAAGTAAATAAGTGTACTAGTACGAAGTCTGTACTGTTGACTAACACC 3555
Db 869 -----
QY 3556 ACGTATGAATATGTGGAGCATATCTCCATGATACAGAAATTTTATGATGTTCAGTCA 3615
Db 876 -----
QY 3616 CAAAATGATTTTGTGATGAAAAAAGTATGATGATGATGATGATGATGATGATGATGAT 3675
Db 877 -----
QY 3676 GATTAACCAAGACGACATGATGTCGTGTGTGTTGTAAGAGTGAAGTGAAGTGAAGTGAAG 3735
Db 889 ValLysAspThrAspArgGlnGly-----
QY 3736 GTACAGTATTAATAAGAAAAAAGCGAAGAAAGATACGAA-----
Db 897 -----
QY 3787 ACAATGATATATATCTTAAGAAAGATGGAAGAAACAATAGATGATGATGATGATGATGAT 3846
Db 916 ThrValLysAsnAlaLeuThrAsnAsnAsp-----
QY 3847 AAAAAGATATGATGA-----
Db 934 LysTyrGlnProGlyGlyLysGlnArgPheProAsnTrpLysCysValSerSerGln 953
QY 3889 ATAAATTTAGT-----
Db 954 LysSerValAlaThrAlaGlySerSerGlyAlaThrGlyLysSerGlyAspLysGlyAla 973
QY 3913 GTGTATGCCCCCTAGACAGCAAAAGTATGCTATCTTGGCAATGATTAAGA 3972
Db 974 IleCysValProProAlaGlyArgArgLeuTyrValGlyLysLeuThrLys-----
QY 3973 ATAAAAAATTAACATCACAAATTAATTAAGAAAGCTTTCATCAATCTGCAGACGA 4032
Db 991 -----
QY 4033 GAACATTTCTTCATGATTAATTAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 4092
Db 997 SerSer-----
QY 4093 AAAGAAATTAAGAGCAAAATCTCCGCAATTTTGTAGATCAATGTTCTACACATTT 4152
Db 1009 Arg-----
QY 4153 GGAGATTATAGATTTTATTTATTTGAACAGATTAATCAAAAGTCAATGATGAGGAGAT 4212
Db 1010 -----
QY 4213 AAATTAAGAGCAATATGATTTCTTTTCAAAAATGTATACCAAAAATCTCTAATGA 4272
Db 1018 -----
QY 4273 AAAACAGCCAGAAATGTGGAGAGAAACATATGATGATATGGAAGTATGATGATGAT 4332
Db 1018 -----
QY 4333 GCACTAGTAAAAATTTGGGGCAAAAAAAGATGATTTTACGAAAAATGAGGTATACAAAC 4392
```

```
Db 1018 ----- 1018
QY 4393 GTCAAATTTAGTACAAAGCACCCTTTGGAGCAATTTGCCAAAGCACCAGTTTAA 4452
Db 1019 -----GlyaspaspIleThrThrGluSerLeuArgLys----- 1030
QY 4453 CGATGGCTAACCGAATGGTACGACGACTATTGCTATACGACAAAATATTGMAAGAT 4512
Db 1031 -----TrpHeiLeu----- 1034
QY 4513 GTGACGAGAAAAATGTAGTCAATGACCAATGTGATGATACAGATTTAATAGAA 4572
Db 1035 -----ThraIaIa 1037
QY 4573 TGGCAGACACGCTTAATATATGAAAAAAGAGTGCATTCACAAAGATTAATAT 4632
Db 1038 ILeGIuThrPheLeuThrPHisArgTylLysIleLysIleuIaIaGlnLysLys--- 1056
QY 4633 TACAAAGATCAACGCCACAAAAAAGATTGATAGACACACATGCGTGTATGTTACA 4692
Db 1056 ----- 1056
QY 4693 GACTATACGTGAACGATGACACAGATTACTGTAGACGAATTTACTGCTAGTTGGT 4752
Db 1056 ----- 1056
QY 4753 GATTAAGCCTCGAAGTCCCTCTGTGTACAAAGAAATATACAAATTTAGAAAAACAGGCT 4812
Db 1057 -----AlaGluLeuGlnArgAsnGlyLeuLeuLeu----- 1066
QY 4813 TACTATGATGCCGACAAACATTTGGGTGCACAAAATTTATTGAAAAATGACCAAAATAT 4872
Db 1066 ----- 1066
QY 4873 ACTAACATTTGAGTAAGTAAGTCAAAAGATTAAGAGGCAACACAGGTGCT 4932
Db 1067 -----GlyThrGlyAla 1070
QY 4933 ATTAAGTGGCAAAACAAAGCTCTTAATACATCAATTAATGAAAGATTAAGTGAAGAT 4992
Db 1071 Ser-----LeuAsnLeuGlnGlyLysAspSerAsnProGlnThrGlnLeuGlnLysSer 1088
QY 4993 GTCCTTTTCCTCTCTGCTGACGATGATATGTTTCATGATGATGATGATGATGATGATGAT 5052
Db 1089 GlyThrIleProLeuAspPheLeuArgIleuMetPheTyrThrLeu---GlyAspTyrArg 1107
QY 5053 GATCCAGAAATTAA-----GATGAAAAATGGGTTCGAAAAAAGATTGATGAA 5100
Db 1108 AspIleLeuValArgGlyValAlaAspAspLysAsnGly----- 1120
QY 5101 GTGGCGCGCAAGGAGGTACAAATTTGGGTCAATTAATCTACAAAGAAAAAAGAGAG 5160
Db 1121 -----GlyAsnAsn----- 1123
QY 5161 AAAATAAAACGTGCGATCGCACAAATATCTTATGAGTCCCGCTGTAGTGTATG 5220
Db 1123 ----- 1123
QY 5221 AAATATAGTTTATGATTAAGATATATTAATCTAGATGATTAATTTGAGAGATGAA 5280
Db 1124 -----IleIleLeuAsnAlaSerGlyAsnLysAspGln 1134
QY 5281 AAACAAGACCGAG-----GAAAATTTGAAGAAAAATTTTAACAAAAATGGAACA 5331
Db 1133 LysGlnLysMetGlnLysIleGlnLysIleGlnLysIleGlnLysIleLeuProThrSerGlyAsn 1154
QY 5333 TCAGTTGGCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5391
Db 1155 LysGlnLysArgGlyProGlnAsnSerValAsnAsp-----ArgGlnSer 1169
QY 5392 TTCTGGAACGAAATTAAGCAATGTGTGTGGAACGCAATGATATGCGGG-----TACAA 5445
Db 1169 ----- 1169

Db 1170 LeuTrpAspArgIleAlaGlnIleValIlePHisGlyMetValCysAlaLeuThrTyrLys 1189
QY 5446 CGTGTAGAGGATGATGGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAGTGAATTAAG 5505
Db 1190 -----AspAspAspAsnGly----- 1194
QY 5506 TGTGTTCTGTACCTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5565
Db 1194 ----- 1194
QY 5566 GCGTATCACTTTCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 5625
Db 1194 ----- 1194
QY 5626 AAGCAATTGAGAAATTTGTTAGGGCGCTGATGATGATGATGATGATGATGATGATGATGAT 5685
Db 1194 ----- 1194
QY 5686 AAAAGAAAGAAATGTACAGATGCTGTACACAAATATTAATAATTTATTAGATGAGAA 5745
Db 1194 ----- 1194
QY 5746 CCACAGTATGAAAAACAAATCAAAAAATATGTTGAGAAATTAATATTCGAG 5805
Db 1194 ----- 1194
QY 5806 CATCGTGTGCAAAAAGATGACAGAGACCGCTCGCAATATTTAGACAAACAAATTAATA 5865
Db 1194 ----- 1194
QY 5866 ATTTGTAAATTAAGAGGAGATTTGAAATTAAGTATGATGATGATGATGATGATGATGAT 5925
Db 1195 -----LeuLysGlyVal----- 1198
QY 5926 CGATTACTGATGATTAATAGTCAAAATATATGCCCCGATCATTTAGACATGAACAAAGAA 5985
Db 1198 ----- 1198
QY 5986 GTTGAAGAAAGTGAATTTGATCAAGTGCACAGAGTCCACACGCTGACGAAGGAAACA 6045
Db 1198 ----- 1198
QY 6046 CCGTACCAAGGGATATCACTGATATCAAAAGCAGCGGATGAAAAAAGACGAAACA 6105
Db 1198 ----- 1198
QY 6106 GCGCGCGCTACAAAAACAGCGAAAAAAGTGGAAATCTTAACAAAGAAATGCGACACAA 6165
Db 1199 -----ValLysLysProGlnLysIleGlnLys----- 1207
QY 6166 ACACGAAACCGACGACGACGACAAACAAACGAAAGACATCAACAGCAACAAACA 6225
Db 1207 ----- 1207
QY 6226 GAATCTGACGTGGCACATGTAAGGCCATTTCTTGAAATAAACAGATAGCAGGGGT 6285
Db 1207 ----- 1207
QY 6286 GGAATAGAGGTTGTAATCCAAAAACGATAGGACATATCTTAATGGGTGTATGTA 6345
Db 1207 ----- 1207
QY 6346 GGTAACTTAAGAAATGAATAATGGCATATGATGCTCTCTAGAGAAAAAATATATGT 6405
Db 1208 -----ProGlnLys----- 1210
QY 6406 ATTAATATATATCAATATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6465
Db 1210 ----- 1210
QY 6466 GAGGCTTTATTAATGTCAGCAATAGAAACATCAATTTTGGTGTAAATAATATATAT 6525
Db 1211 -----LeuTrp----- 1212
```


Db 1586 LysThrSerAspGluTyrPheAsnValIysSerPheLeuGluThrTrpIleProLys--- 1604
QY 8680 TATAAAGAGGCTTCACCTTGTACACTACCTTCAAGATCCATTGGATTAATGTCTTACC 8739
Db 1604 ----- 1604
QY 8740 GATCAAAACAAAGATGATGTAAAGAAATTAACAACCTTTACCTTCCTGCAAGAAATGAT 8799
Db 1605 ----- 1605
QY 8800 TATGATTAATATCTTGTATTAATGGAACGCAATACCTTGTCTTAAATAGTTGATGATATAC 8859
Db 1611 GlnAspAsn-----ValIleLysLeuSerLysPheGlyAsnSer 1623
QY 8860 AAGGTGTATGATCTCTCCAGAGAGAGACATTTATGTACAGACCTATACATGCAATAT 8919
Db 1624 CysGly-----CysSerAlaSerAlaIleSerThr 1633
QY 8920 AATTATAGAAAGGTGATTAAGAAATTTTAAAAAAACCTTCTACTCTGCTTTCAGT 8979
Db 1634 Asn-----GlyAsnGluGlu----- 1638
QY 8980 CAAGGCAATTTGTAGGTCAAAAATTAATCGGAAGAGAGTTGCTTTGAGGCAATG 9039
Db 1638 ----- 1638
QY 9040 AATATAGTATGACAGATTAATTCGATTAATTAAGAACTGATATGATGACACTTCA 9099
Db 1639 ----- 1639
QY 9100 TTTATCGAAAAATTAATAAAATATTTTGAACACTCAATGAAGCAACGAAATCTGTA 9159
Db 1645 IleLysLysLeuGluLysLysLysIleAspGluCysLysArgLysProGlyLysAsn 1662
QY 9160 ACATGTGGGAAAAATATATAGACGTGATATGCGACGTATGTATGTGATATTAAT 9219
Db 1662 ----- 1662
QY 9220 GCTACTTCAAAAGTAACTAGATAGTAGAGATGGTCAATTAACAAGAGATGAGAACT 9279
Db 1662 ----- 1662
QY 9280 AATGAGTTCTTCTGTTGTTAATGAATGGCAAGCAAGCATGTATAGGAAAGAACT 9339
Db 1663 ----- 1663
QY 9340 GTAAGTATTCATTAATAAACAATAATGCTCCTGCAACGAGATTAATTTGAGCGTCA 9399
Db 1668 ----- 1668
QY 9400 GAATATTTAAGACACCTGATGTGATGATATTAAGAAATTAATTAAGCTTGAATATA 9459
Db 1669 GluThrLeuThrHisPro----- 1674
QY 9460 TTGATAAAAAATACATGGAATAATCTAATATATAAATAATTAAGCATTAAGATCAATCT 9519
Db 1674 ----- 1674
QY 9520 TCAGGTAATATAGACAAATTAACCATCTGAGAGAAATGTTCACTCATATATAAATCAAAA 9579
Db 1674 ----- 1674
QY 9580 GATTCATATGCGCTTGGAGTTAATATATTAATGAATAGTTACAGAAACAAAT 9639
Db 1675 ----- 1675
QY 9640 AATGAATAATATGATCAAGAAGTACTAAAAAAATTAATATCCGCTTATATAT---TTT 9696
Db 1690 AsnPro-----ValGlyLysGlnHisProSerPheCysProPro 1702
QY 9697 GTTGAAGTGAACACACAAATAATCAGTACTAGATGGAATATATAAAGAGAGAG--- 9753
Db 1702 ----- 1702

Db 1703 ValGluAspLysLys-----LysGluGluGluGly 1712
QY 9754 CAACAGTTGTCCTTAAGCACTATTTCTTACACCCCATGATGATCTTTATCA 9813
Db 1713 GluThrCysThrProLaser-----ProLaserProLaserProLapro 1726
QY 9814 GCACCTTTATTCACACATGACATGACAGACATATGATCTTAATAAATAT---ATA 9867
Db 1727 AlaProLaserProSerProThrProLaproLaserGluProPheAspProThrIle 1746
QY 9868 TTGAAGATAGTATCTCTGTATTTATGATCGGCTTGGTGTATGATGAGCTTATTC 9927
Db 1747 LeuGlnThrThrIleProLeuGlyIleAlaLeuAlaLeuGlySerIleAlaPheLeuPhe 1766
QY 9928 ATGAG----- 9933
Db 1767 LeuLysValIleTyrIleCysValValTyrMetTyrIleTyrMetCysPheCysIleTyr 1786
QY 9934 -----AAAAATTCAAATCGTCTGAGCTTGTGCGATATACATATCCGCAAGA 9987
Db 1787 MetTyrLysLysThrLysHisProValAspLeuPheSerValIleAsnIleProLysSer 1806
QY 9988 GAGTATGAAATCGCTACGTTGGATCCAAATAATAGTACATATACATATAGAGTGTCCA 10047
Db 1807 AspTyrAspIleProThrLysLeuSerProAsnArgTyrIleProTyrThrSeriLys 1826
QY 10048 TTTAAAGCAAAACATATATATATATGAGAGAGATCT-----AGTGAAGATCA 10098
Db 1827 TyrArgGlyLysArgTyrIleTyrLeuGluLysAspSerGlyThrAspSerGlyThr 1846
QY 10099 GATTAATATATGATGAGATTAATCTCTGATATATCT---TATCGGAAGAGATAT 10155
Db 1847 AsnHisTyr-----SerAspIleThrSerSerSerGluSerGluTyr 1860
QY 10156 GAAGATGATGAT 10215
Db 1861 GluGluMetAspIleAsnAspIleTyrValProGlySerProLysLysThrLeuIle 1880
QY 10216 GAAGTATGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 10269
Db 1881 GluValValIleGluProSerGlyAsnAsnThrThrAlaSerAspThrGlnAsnAspIle 1900
QY 10270 ---AATGATACACACGATGATGATGATGATGATGATGATGATGATGATGATGAT 10326
Db 1901 GlnAsnAspGlyIleProSerAsnLysPheSerAspAsnGluThrPheLeuLysAsp 1920
QY 10327 GATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10380
Db 1921 AsnPheIleSerAsnMetLeuGlnAsnGlnProLysAspValProAsnAspTyrLysSer 1940
QY 10381 GCAGATATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10440
Db 1941 GlyAspIleProPheAsnThrGlnProAsnThrLeuTyrPheAspLysProGluGluLys 1960
QY 10441 CTTTAT 10500
Db 1961 ProPheIleThrSerIleHisAspArgAsnLeuLeuAsnGluGluTyrSerTyrAsn 1980
QY 10501 ATTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10560
Db 1981 ValAsnMetSerThrAsnSerMetAspAsp---ProLysTyrValSerAsnAsnValTyr 1999
QY 10561 AGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 10619
Db 2000 SerCylIleAspLeuIleAsnAspSerLeuSerCylAsnLysHisIleAspIleTyrAsp 2019
QY 10620 GAACATATG 10628
Db 2020 GluValLeu 2022
RESULT 9
AAW22476
ID AAW22476 standard; Protein; 2182 AA.

OY	3163	ACCCAAATGGCGAAGATGTACTCTCGAAGGTGCAGAAAAAGAGATATGATTAAGTTGAAAGG	3222
Db	844	 Thrlu-----GlmU	847
OY	3223	AAGGTAAAGAGTGTAAAGATTAAGTAAATGTCTCAAGGCTGTACGAAGAGTGGTACA	3282
Db	848	 GlyVallylprocy-----	852
OY	3283	GGTTGTACGAAGTGCACAGAACTTGTAAATGAATATAAATGATATATAGATTATAGCAA	3342
Db	853	----- AspIleValGlyLysLeu---	858
OY	3343	GAACAAATGAATATTAATATCATGATTAATACAAAAGAAATACATGAACAACAAATGCT	3402
Db	858	-----	858
OY	3403	GTTAGTAATAGTGGATTGAAAGCTTCAGTACTGCCAAAATCATATAGACAGAAATGT	3462
Db	858	-----	858
OY	3463	ATTGAATTTTTCGCGAATTATACCACAAAATGTGGCAAAAGTAATAAAAGTGTACT	3522
Db	858	-----	858
OY	3523	AGTGATGAAGTGTGCTGCTATTGGTACTAACACACGATGAATAAGTTGAGCATATCTC	3582
Db	858	-----	858
OY	3583	CATGATACAGMAATTTTGATGATGTGCAGTCACAAATAGTTTGTATGAAAAAGT	3642
Db	859	 -----PheGluAspAllySerLeuGlu---	867
OY	3643	GATGTAAAGATTAACGAAAAATATGCTTTAGAGATAAACACAGACCATGATGTGCG	3702
Db	868	----- Ala	868
OY	3703	TGTGTTGTAAAGTGTGATCGAAACCGACAGAGGTACAGATAAAAAGAAAAAGCG	3762
Db	869	 CysGlyLeuLysTyrglyPro-----	875
OY	3763	GAAGAAAAGATACGGAATTAACACGATGATATCTTAAAGAAAAGCATGGAAG	3822
Db	876	----- GlyGlyLys	878
OY	3823	AAACAATAGAAAGTGTCTCATCCAAAAAGAAATAGTAATGCCATGCCATGGCAATGC	3882
Db	879	 GluLys-----PheProAsnTrpLysCys	886
OY	3883	GAAGATTAATATTAGT-----GAAGACCCCTCGTGTGTATGGCC	3924
Db	887	ValhrProSerGlyValSerThrAlaThrSerGlyLysAspIleAlaIleCysValPro	906
OY	3925	CTTAGAAGACAAAAGTTATCGTACATTTCTTGCA-----AATGATAT	3969
Db	907	ProAlrGlyAlrGlyLeuTyrlValGlyGlyLeuSerGlnTrpAlaSerArgIlyLysp	926
OY	3970	GAATTAATAAATATACAAATCAAGT-----AAT	3999
Db	927	GluThrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGluLys	946
OY	4000	TTAAAGAAAGCTTCATCAATATCGACGACGAACAATCTTCTCATGATGATATAT	4059
Db	947	LeuArgThrAlaPheIleGluSerAlaAlaIleGluThrPhePheLeuTrpHisLysTr	966
OY	4060	AAA-----ACTAAGATGT-----	4074
Db	967	LysGlnLysLysProAlaIleThrGlnAspGlyAlaGlyLeuGlyValSerLeuPro	986
OY	4075	-----GAAGCAATGAACCTCGATTAAGATTAATAAGAA-----GGCAAAATTCCT	4119
Db	987	GluProSerProProGlyGluAspProGlnThrGlnLeuGlnGlnThrGlyValIlePro	1006

OY	4120	CCCCATTTTGAAGATTCACAGTCTACACATTTGGAGATTATAGATTTTATTTGGA	4119
Dd	1007		1024
OY	4180	Proaspheuleuarvglmethpetyrlhleualasptyllyasplileu-----	1024
OY	4180	ACAGATATATCAGAAAAGGTCAATGGTGAGGGAAGTAAACTTAAAGCAAAATAGATTCCTT	4239
Dd	1024	-----	1024
OY	4240	TTCCAAAATGGTGACCAAAAATCTCCTAATGGAAAAACAGCCAGAAATGGTGACAGAA	4299
Dd	1024	-----	1024
OY	4300	CATAGTCATGAGATATGGAAAGCTATGCTATGTCGACTGTAAAAATTGGGGCAAAAAA	4359
Dd	1024	-----	1024
OY	4360	GATGATTTTACCGAATACTACGGTTACAAACAGTCACAAATTTAGTACAAAAAGCACACT	4419
Dd	1024	-----	1024
OY	4420	TTGGAGGAATTTGCCAAACGACCCAGTTTTCATGCTAGGCTAACCGAAATGGTACAGACAC	4479
Dd	1024	-----	1024
OY	4480	TATTCGTATACAGCAAAAAATTTTGAAGAGATGTGCAGAAAAATGTAAAGCAAAATGAC	4539
Dd	1024	-----	1024
OY	4540	CAATTTGAAGTGTGATACAGAAATGTAAATAGAAATCGAGACTACGTTAAATATATGAAA	4599
Dd	1024	-----	1024
OY	4600	AAAAAAAAGAGTGGATTTCCACAAGATAAATATTCACAGATGAACGGACAAAAAAGA	4659
Dd	1024	-----	1024
OY	4660	TTTCATAGACACACATTTGGTGTAAATGTTTACAGACTATCTAGCGAAGCATATGCAAGAT	4719
Dd	1025	-----	1032
OY	4720	TACTTGAACAGAAATTTACTGCTAGTGTGTGATGAACCTGGAAGTGCCTGTGTGTA	4779
Dd	1032	-----	1032
OY	4780	CAAGAAATATACATATTTGTAAGAAAAACAGCTTACTATGATCCGACAAACATTTGGG	4839
Dd	1032	-----	1032
OY	4840	TGCACAAAATTTATTTGAATTTGACGACAAATATACTACATTTTCGAGTAAAGATTAAGTC	4899
Dd	1033	-----	1047
OY	4900	AAAGATTTAGTAAAGAGGCAACACACAGGTCTATTAAGTGCAAAACAAAGGTCTTAT	4959
Dd	1048	LysasnileYallleugluYalaser-----	1055
OY	4960	AACTACATATACCTTGAAGAAGATTGACTGAGATGTCTTTTCTCTCTGTCGACTACGT	5019
Dd	1055	-----	1055
OY	5020	ATATGTTTTCATGCTATGGATGGCAATTTATACAGATCCAGAAATTAAAGATGAATAATGGG	5079
Dd	1055	-----	1055
OY	5080	TTGGGAAAAAGATTGATGGAAGTGGGGCAACGGAAGGGTTACAAATTTGGTCAATACTAC	5139
Dd	1056	-----	1059
OY	5140	AAAAAAGAAAAAGAAAAAGAAAAATAAAAACCTCGATGCCACAAATATCTTATGAG	5199
Dd	1060	-----	1066
OY	5200	GTCCGCGCTTGTAGTGTATGAATATATAGTTTTTATGATTTAAGAGATTAATTTCTAGT	5259

```

Db 1066 ----- 1066
QY 5260 ATTGATATTTGGAAGATGAAAAACAAACGAGGAAAAATTGACAAATATTTAAC 5319
Db 1067 ----- 1077
QY 5320 AAAAAATGACATCAGTTGGCAAGAGATGATACAGGAAATCCCGTAGTACT 5379
Db 1078 ---GlyAlaThrSer-----GlyValProValThrLysAsnSerValLysThr 1093
QY 5380 GCGGAAATTTTCTGGAAGAAAAATAGAAATGTGTGGAAACGATGATATGCGGG 5439
Db 1094 ProGlnInhrTrpTrpGluAsnIleAlaLysAspIleTrpAsnIleMetValCysAla 1113
QY 5440 -----TACAACGTGTGTAGGATGATGAATAGTGAATAGTGCACAGATGATGA 5493
Db 1114 LeuThrLysGlu-----AsnAspAlaTrp----- 1122
QY 5494 GATCTAAAAAATGTGTTCTGTACTCTCAGATGATGATTTCTTATGGGAAAAATCGC 5553
Db 1122 ----- 1122
QY 5554 GATGAAGTACTGCTATCAGTTTCTTCGATGGTTTCCGCAATGGGCTAAGATTTTTCG 5613
Db 1123 ---GlyThrSer----- 1125
QY 5614 AAACATTAAGAAAAAGAAATTGAGAAATTCGTAGGGCGCTGAATGATTATCTGTGGT 5673
Db 1125 ----- 1125
QY 5674 GATAATGAAGTAAAAAGAAATGTACAGATGCGCTGACACATATATAAAATTTAAT 5733
Db 1125 ----- 1125
QY 5734 AGTAGTGGAAACACAGATGAAAAAACAATCAAAATATGTGAGATTAAGACAAA 5793
Db 1126 -----AlaLysIleGluInsnLysAsp--- 1133
QY 5794 ATATATTCGAGATCCTGTGGCAAAAGATGCAGAGACGCTGGCAATATTTAGACAAA 5853
Db 1133 ----- 1133
QY 5854 CAAATTAATAATTTGTGAATAATAAGTGGAGATTTGATATTAAGTATGAAGAT 5913
Db 1134 ---LeuLysLys----- 1136
QY 5914 GTGTCCACACAGCATTAAGTATGATGATTAAGTCAAAATATGCGCATATTAGACAT 5973
Db 1137 ---AlaLeuTrpAspGluAlaAsnLysAsnThrPro----- 1147
QY 5974 GAACGAAAAAGATTTGAAGAAAGTGTATTTGTCAAGTCCACAGAGTCCACCAGCTGT 6033
Db 1147 ----- 1147
QY 6034 CGAAGGAAACACCGCTACACAGGGTATCAGTATCAAAAGCGACGATCGAAAAA 6093
Db 1147 ----- 1147
QY 6094 GAAGGAAACACGCGCGCTTACAAACAGCCGAAAAAAGTGGAAATCTAACACAGAA 6153
Db 1147 ----- 1147
QY 6154 ATGCGACACAAACACGAAACCCGAGCAGACAGACAAACAAACGAAACGATCAACA 6213
Db 1147 ----- 1147
QY 6214 GCAACACACAGAAATCTGACGTGGCACAATGTAAGGCCATCTTTGCAATAACCA 6273
Db 1147 ----- 1147
QY 6274 GATAGCAGGGGTGAATGAGGGTTGTATCCAAAAACGTATGACATATCTTAATGG 6333

```

```

Db 1147 ----- 1147
QY 6334 GGTGTATTTGTAGTAAGTCTAAAGAAAAATGAAATGCGATGATGCTCTAGAGA 6393
Db 1147 ----- 1147
QY 6394 AAAAAATTTGATTAATAATATACAAATTTTAATATGAACCTGAAATTAAGGTGAC 6453
Db 1148 ---IleGluLysTrpGluTrpAsnValLysLeuGlu----- 1159
QY 6454 AATGATTAAGAGAGCTTTTATTAATGTGACAGAAATGAAATTCATTTTGTGTTA 6513
Db 1159 ----- 1159
QY 6514 AATATATATTTGAAAAATCCCTGACACGAAAAATGAATTCACAAATTCAGAT 6573
Db 1159 ----- 1159
QY 6574 GAATTTAAAGAAATATGATATATACATATGTGATTTAAAGATATGTTTGGAACT 6633
Db 1159 ----- 1159
QY 6634 GATATTTTAATGATAAAAAATTAATTAACGTGAACAATAGTGAACAACCATTCGAT 6693
Db 1160 -----Asp 1160
QY 6694 GAAATTAATAGAAAAACAGATTAATAAAGATTAAGTAATTAATTTTGG 6753
Db 1161 GluSerGlyAlaLysSerAsnAsp----- 1168
QY 6754 GAGAAAAATTAATAATTTATTTGGAGAGATGATATATGATTAACTTATCATCTACA 6813
Db 1168 ----- 1168
QY 6814 GACGAAAAAGAAAAAATTAAGATTAATTAACAGTACATGACATGCCAACTG 6873
Db 1169 -----ThrIleGln 1171
QY 6874 ACGCTTCCCTTGAAGATTTGTAAAAAGGCCCAATTTTGAATGTTTACACGAATGG 6933
Db 1172 ProProthrLeuLysAsnRheValGluIleProthrPhePheArgTrpLeuInsnLysTrp 1191
QY 6934 GCAGAGAAATTTTGTAAATAGAGAGAACAGTTGTTAAATGTGAGCGGGCTGT 6990
Db 1192 GlyAsnSerPheCysPheGluArgAlaLysArgLeuAlaGlnIleLysHisGluCysMet 1211
QY 6991 ---AAGAAATATGAGTGTATGCTGTGTAATGACGGAAG----- 7026
Db 1212 AspGluAspGlyGluLysGlnTrp-----SerGlyAspGlyGluTrpCysGlu 1227
QY 7027 -----ACACAAGAAATGGCAGAGCGG 7047
Db 1228 GluIlePheSerLysGlnTrpAsnValLeuGlnAspLeuSerSerCysAlaLysPro 1247
QY 7048 TGTGTAACATATCAAAATTTTATTAAGAGTGGAAACGTAATATGAACAAAGAA 7107
Db 1248 CysArgLeuTrpLysTrpIleGlnLysLysThrGlnLysGlnLys 1267
QY 7108 AAGTTCAAAAAGATTAAGATGCAAAAAATATAGATATATCTTACTGAAAAGAC 7167
Db 1268 AlaTrpGluGlnLys-----SerAsnTrpLysGlnLysAsp 1282
QY 7168 ATAGAGAGCAACATGTGCTCATGATATTTAAACATGAATTAAGAATATATGTGC 7227
Db 1282 ----- 1282
QY 7228 AATAGGATTTGTTCTGTATGCAAAACCTTTTCAACACTACCAAAACACACACAA 7287
Db 1283 ---LysCys-----GlnThr 1286
QY 7288 TCACATCATCCGATGCTATGATGATGCCAGATGCGTGGATTAATGCTCT-----GAA 7341
Db 1287 GluSerAsnAsnAsnAlaAsnGlnPheSerArgThrLeuGlyAlaSerProThrAlaAla 1306

```

QY 7342 GAATTTACAAAGTGTGAGTGTCTGAACTTTCAAAAAAGGAGTCTATGATTCATACAAA 7401
|||
Db 1307 GluPhe-----LeuGlnLysLeuLysSer----- 1314
QY 7402 AAATTTACGAACTTAATACCTATGATTTGTAGAGAACAGCATATTAATTTATCT 7461
|||
Db 1315 -----CysLysAsnAspAsnGlyTyr----- 1321
QY 7462 AAAGAACGAAATAATATGATATTAATACCTTGAAGAAAAATTAATACCTATGAGTCT 7521
|||
Db 1322 -----GluAsnGlyGluAspAsnLysLysLeuAspPheLys-----AsnProAspLysThr 1337
QY 7522 ACAAGAGAAAAAGAAAGTAAATAGTTGAGCTAATTAATCTTGCATCCATAGAAA 7581
|||
Db 1338 PheLysGluLysLysSer-----CysAspPro-----Cys 1347
QY 7582 CCTATACACCTGATTAATATATAGAGAGAAAGAACCTTGTGAAATATAGAGAGAAAT 7641
|||
Db 1348 PheLysGlnLysLysSer-----CysGlnAsnGlyHis----- 1358
QY 7642 CGTTTAAAGTAGATTAATGATGGAATGTTACAAAAATTCAAAGTCTATCAGAGAAA 7701
|||
Db 1358 ----- 1358
QY 7702 AAAAGAGTATGTGTACCTCCAGAGAGAACATATGTGCTTAAGCAATTTAGATGAAT 7761
|||
Db 1359 -----CysValGlySerAlaAsnGlyLysGlnCysLysAsnAsn-----LysIle 1373
QY 7762 AAATTTGAAGACTTAAGATAGTAAATTAATCTCTAATAATGCTTCCAGACGCGACGA 7821
|||
Db 1374 ThrAlaGluAspLysLysAsn-----LysThrAspPro 1384
QY 7822 AATGAAGAGTATGACATATAATAAACTCACTCAGAGAGACGGGTGCGCAATGAATCA 7881
|||
Db 1385 AsnGlyAsnLysGlnMetValLysSerAspSerThrAsn----- 1398
QY 7882 ATATGATATCTATGAATATATAGTTTCTGCTGATGCTGGTGCATAGTTAGAGAACAGAT 7941
|||
Db 1399 -----ThrPheGlnHisLeuGlyAsp----- 1405
QY 7942 ATGTTAGCAATGTGTGTACTTACTCCGCTAGCAATTAATATATTAAGCTTTTGAA 8001
|||
Db 1406 -----CysLysSerSerGlyLysPheLys 1413
QY 8002 TACATATGTAAGAAATGAGAAATTAATAAGTAGAATAATAACAAGATGTACAA 8061
|||
Db 1414 GlyIle----- 1415
QY 8062 ACGTTTGTCTGCTGTGGGATGCTAATAGAAAGATATTTGGAAA-----GCAATGAGC 8118
|||
Db 1416 -----ArgLysAspGlnTyrPlyScyAlaAsnVal 1425
QY 8119 TGCAAACACCAGAACATGCAAAACTTTTGAAGAAAGAAATGATGCAATTTGAAAGC 8178
|||
Db 1426 Cys-----GlyValAspIle 1430
QY 8179 ATACATTAAATACAAAT-----AAGTGTGACATTAAGAGACGATCCAGCTGTGATGATAT 8235
|||
Db 1431 CysThrLeuGlnLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1447
QY 8236 ATACCTCAAGCTTTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 8295
|||
Db 1448 IleThrMetLys-----GluLeuLeuLysArgTyrPheLysLysLysLysLysLysLys 1462
QY 8296 GAAGAATTTGAAAAATTTAAAAATTCATGTGATCAGTCT-----AAAATCCTGACAGATGC 8332
|||
Db 1463 GluAspLysAsnArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys 1482
QY 8353 AAGATGATTAATGATGAATAATAGTGAACAGGTAAACGATGTCAAGATATATAA 8412
|||
Db 1483 Lys-----CysIleLysGlyCysLysIleGln----- 1490

QY 8413 AATTTTGTCTTAATGGAATCTCTATTTGATATACATCAAAATTAATTAACAAAGATTG 8472
|||
Db 1491 -----LysTrp----- 1492
QY 8473 TATGAACAACCAATATATACAAAATCTACTATATGATCATGTCAAAATTTGTACAA 8532
|||
Db 1493 -----ValGln-----Glu 1495
QY 8533 AAGTGAATACTTTTAAAGATGATGTTCTGTGAGAGCTTTCTGAAATCTTCATGAA 8592
|||
Db 1496 LysThrLysGlnTyrPheLys----- 1501
QY 8593 ACAAGTAAGTGTGATTAATTAATTAATGAAGAAATGATGTTCTTCCATATAGCAACA 8652
|||
Db 1502 -----LysIleAsnAsp-----Thr 1506
QY 8653 TATGCTTCGAGAGAACCAACAAAAGTATTAAGAAAGCTTGACAGTGTACATACCTCT 8712
|||
Db 1507 Tyr----- 1507
QY 8713 AAGATTCATTGATTAATGTCTACCGATCAAAACAAAGATGATGATGAATTAACA 8772
|||
Db 1508 -----LeuGln 1509
QY 8773 ACTTTTACCTCTGCTCGAAGATGATTAATTAATATATCTGATATTTGAAACGATAC 8832
|||
Db 1510 GlnTyr-----LysAsnAspAspGlyAsnThrLeuThrAsnPhe----- 1522
QY 8833 CTGTCTTAAATAGTTCAGATGATTAACAAGAGTATGATTTCTCTCCAGAGAAAGCAT 8892
|||
Db 1522 ----- 1522
QY 8893 TTATGTACAAGACCTATACGTGATATTAATTAAGAAAGGTATTAAGAAATTTTAAA 8952
|||
Db 1523 -----LeuGlnGlnPheGlnTyrAla----- 1529
QY 8953 AAAAACTTCTTACTTCTGCTTCTTACGCAAGACAAATTTAGGTCAAAAATTAATTAATCG 9012
|||
Db 1529 ----- 1529
QY 9013 GAAGAAAGTGTGTCTTTGAGGCAATGAATTAATTAATGATATGCAATTAATTCGATTAAT 9072
|||
Db 1530 -----ThrGlnPheLysAsnAlaIle 1536
QY 9073 AAAGAACTGATATGATGACACTTCATTAATCTGAATAAAATTAATAAAATATTTGAACA 9132
|||
Db 1537 LysProCysAspGlyLeuAsp----- 1543
QY 9133 TCAATGAAGCAACCGAAAAATCGTAAACATGTGTGGAAAAATATATAGCGTCAAGATGCG 9192
|||
Db 1543 ----- 1543
QY 9193 CAGCGTATGTTATGTGATTAATAATTTGCTACTTCAAAAGTAACATTATGATGAAGATGCG 9252
|||
Db 1546 ----- 1546
QY 9253 TGTCATTACCAAGATGAAGAAATAATCAAGTTTCTGTTGCTTAATTAATGAATGGCA 9312
|||
Db 1544 -----GlnPhe----- 1545
QY 9313 AAGCAACGATGAAGAAAAAGAAACATGATGATGATTAATAAAACAAATGCTCTCGT 9372
|||
Db 1546 -----LysThrSerGlyLeu 1551
QY 9373 TCAACGAAATTAATTTTGAAGCGTCAAGATTAATTAAGACAACTGATGTCAGATGAT 9432
|||
Db 1552 AsnSerThrAspAsn-----SerGlnAsnGlyAsnAsnAsnAsp 1564
QY 9433 ATTAAGAAATATATATTAATGATTAATTAATAAAATCAATGGAATAATTAATATA 9492
|||
Db 1565 -----LeuValLeuCysLeuLeuAsnLysLeuGlnLys 1575
QY 9493 AAATATTAAGCAATTAATAAGATCAATCTTCAAGTAAATATAGACAAATAAACCATCTGAAGA 9552

QY 2005 GAAGGTATTTTAAAGTTATGATTAACCTTGACAAAGATGAAAGCAAAATGAGAAAGAA 2064
Db 691 ValPheSerProTyrGlyValLeuAspLeuVal-----LeuYsGlyGlyAsn 706
QY 2065 CTATATGGAATAATTAATAAATAAATAAATGACATTTTCCATTTGGAAATAATAGGAC 2124
Db 707 LeuLeuGlnAsnIleYsAspValHisGlyAspThrAspIleYsHisIleYs 726
QY 2125 TATTTAGAG-----AATGCAATA 2142
Db 727 LeuLeuAspGluGluAspAlaValAlaValLeuGlyGlyYsAspAsnThrIle 746
QY 2143 GAACCTCTGTAGATCACTTAAGAAAGAACTGCCAGATATGAAGACATAATACAAAC 2202
Db 747 AspYsLeuLeuGlnHisGlyGluGlnHisGlyGluGlnHisGlyGlnHisGln 764
QY 2203 GAAGCAATGTGAACATCCCATTAATGCAACAACAACCCGCTGTAACTCTGTGAGGC 2262
Db 765 GluGluCysGlu----- 768
QY 2263 ACGCAACCCACTTAATAATTAATAAAGAAATAGCAAAATACCTTAAAGAGATGATACGAG 2322
Db 769 -----LysYsAlaGlnGln 773
QY 2323 GAAGCAGAAATCGTGCTCTCTAATTTGAAGAAGAAAGCAGAGATATATATAA 2382
Db 774 GluSerArgGlyArgSerAlaGlu----- 781
QY 2383 CGTGCGGTAGAGAAAGCACTTCAAGACAAATTTATGATATATGATTAACATTCCT 2442
Db 782 -----ThrArgGluAspGluArg----- 787
QY 2443 AATCGTAATCTGGTTTTCAAATGACCATGTGATGCGAAAGCAGCATGATGTATA 2502
Db 788 -----ThrGlnGlnProAlaAspSerAlaGlyGlu----- 797
QY 2503 CAACAAGATTTGTCTAGAACTGATGGAAGTGAAGTCCGGAACACATGCGTAAAGT 2562
Db 798 -----ValGluGluGluGluAspAspAspAspPyrAspGlu 810
QY 2563 CACGAGATGTTATTATGCTCTCTAGAAAGCAGCATATGTACATCCAAATTTGAAAT 2622
Db 811 AspGluAsp----- 813
QY 2623 TTACAAGCGATGATCACCCACTTAATGTAATATTTGATGATTTAGTAAATATCC 2682
Db 814 -----AspAspValValGlnGluGlu 820
QY 2683 TTTTGGGGATGTTCTCTATCAGCAAAATATGAAAGCAACAAGATATACGATGAT 2742
Db 821 GluGluGly----- 823
QY 2743 AAAGAAAGATTAACCTTAAGGGCCCCAAAGATTAAGTGAACCCAAACACGAGACAAT 2802
Db 824 LysGluGluGluGlyThrValThrGluValThrGluValThrGluValGluGluThr 842
QY 2803 ATCTGTGAGCATATCGTTACGTTTTCAGATATAGGTATATATTCGAGGAAGAT 2862
Db 842 ----- 842
QY 2863 CTCTGGAAAGAAACGATGACATGTAAGCTGCAAGACATTTGAAACTGTTTGT 2922
Db 842 ----- 842
QY 2923 AATATACATTAAGTCACTCAAGAGCAAGAAAGATTAATATATGATGATCCCCCAA 2982
Db 842 ----- 842
QY 2983 TATTTAAATTTGAGGAAATTTGGTGGAGACTAATAGCCAAAGTATGGAAACCATG 3042
Db 842 ----- 842
QY 3043 AAATGTGATATAAATATTTGAGAGATTAATCGGAGACCAATCAACAAAGTATAT 3102
Db 842 ----- 842
QY 3103 TCGGATATAGTATGATCATACACCATTTGATGATTAATCCCAAAAATTAAGATGATG 3162
Db 843 -----Val 843
QY 3163 ACCGAATGGCAGAAATGTACTGCAAGGTGCAGAAAAGAGATATAGTGAAGAG 3222
Db 844 ThrGlu-----GlnGlu 847
QY 3223 AAGTGAAGAGTGAAGATTAAGATTAATGCTCAAGGCTGTACGAAGAGTGTACA 3282
Db 848 GlyValLysProCys----- 852
QY 3283 GGTGTACAGAGTGCACAGAAAGCTTTAATGATTAATTAATTAATGATTAATGAAA 3342
Db 853 -----AspIleValGlyLysLeu 858
QY 3343 GAACAATGGAATATATATCATGATTAATATCAAGAAATTCATGAACAAGCAAAATGCT 3402
Db 858 ----- 858
QY 3403 GTTATTAATAGTGTATTTGAAGCTTCAGTACTGCCAAAATCATATAGACAGAAATGT 3462
Db 858 ----- 858
QY 3463 ATTGAATTTTGTGCGAATTTTACCAACAATAAGTGTGCAAAAGTAAATGAAGTACT 3522
Db 858 ----- 858
QY 3523 AGTGATGAAGTGTCTGATGTGATGATCAACAACAAGATATGAATGTTGAGCATATCTC 3582
Db 858 ----- 858
QY 3583 CATGATACAGAAATTTGATGATGTGCTGACACAAATATGATTTGATGAATAAAGT 3642
Db 859 -----PheGluAspAspLysSerLeuYsGlu 867
QY 3643 GATGTAAGATTAACGAAAAATATGCTTTAGACATAAACACAGACCATGATGTCG 3702
Db 868 -----Ala 868
QY 3703 TGTGTTGTAAGATGCTGATGTCGAAACGACAGAGGTACAGATTAATAAGAAAAAAGCG 3762
Db 869 CysGlyLeuYsTyrGlyPro----- 875
QY 3763 GAAGAAAGATACGAAATGTAAACAGTGAATATACTTAAGAAAAAGATGAAG 3822
Db 876 -----GlyGlyLys 878
QY 3823 AAACAAGTAAAGATGTGATCCAAAAAAGATTAAGTATCCCATGTCGAATGC 3882
Db 879 GluYs-----PheProAsnTrpLysCys 886
QY 3883 GGAATATTAATTAATG- - - - -GAAAGCCCTGCTGTATGCCC 3924
Db 887 ValThrProSerGlyValSerThrAlaThrSerGlyYsAspGlyAlaIleCysValPro 906
QY 3925 CCTGAAGACAAAAGTATGCTGATCTTCTGCA-----AATGATAT 3969
Db 907 ProArgArgArgLeuTyrValGlyLeuSerGlnTrpAlaSerArgGlyLys 926
QY 3970 GAAATTAATAAATTAACAATCAAGTT-----AAT 3999
Db 927 GluThrThrGluValSerSerGluAlaThrSerAlaProSerGlnSerGluYs 946
QY 4000 TTAAGAAAGCTTTCATCAATCTGCAGCAGCAAGAAACATTTCTCATAGTATATAT 4059
Db 947 LeuArgThrAlaPheIleGluSerAlaAlaIleGluThrPhePheLeuTrpHisYsTyr 966
QY 4060 AAA-----AGTAAAGATGT----- 4074


```
Db 967 LysGluLutLysLysProProLathrGlnAspGlyAlaGlyLeuGlyValSerLeuPro 986
QY 4075 -----GAGGAATGAACCTCGATTAAGAAATTAAGAA---GGCAAAATTCCT 4119
Db 987 GluProSerProProGlyGlnAspGlnThrGlnLeuGlnGlnThrGlyValIlePro 1006
QY 4120 CCCGATTTTGGATGCCATGTTCTACATTTGGAGATTATAGAGTTTATTATTGA 4179
Db 1007 ProAspPheLeuArgGlnMetPheTyrThrLeuAlaAspTyrLysAspIleLeu----- 1024
QY 4180 ACAGATATATCAAAAGGCTATGCTGAGGGAAGTAACATAAAGCAAAATAGATTCTCT 4239
Db 1024 ----- 1024
QY 4240 TTCAAAATGTGACCAAAAATCTCCTAATGAAACACGCCAAGATGTGACAGAA 4299
Db 1024 ----- 1024
QY 4300 CATAGTCATGAGATATGGGAAGCTATGCTATGTGCTACTATAAAATTTGGGGCAAAAA 4359
Db 1024 ----- 1024
QY 4360 GATGATTTTACGAAACCTACGTTTCAACAACAGCTCAAAATTTAGTACAAAAGACACT 4419
Db 1024 ----- 1024
QY 4420 TTGGAGAAATTTGCCAAACGACCCAGTTTTCAGATGGCTAACGAATGTACAGAC 4479
Db 1024 ----- 1024
QY 4480 TATTGCTATACAGACAAAAATATTGAAGAGTGTGACGAAAAATGTAACTAATGAC 4539
Db 1024 ----- 1024
QY 4540 CAATTGAAGTGTATACAGATGTAATAGAATGCGAGACTACGTTAAATATATGAAA 4599
Db 1024 ----- 1024
QY 4600 AAAAAAAGATGTGATTCACACAAGATTAATATTACAAAGTGAACGCCAACAAGAA 4659
Db 1024 ----- 1024
QY 4660 TTGATAGACACACATTTGTTAATGTTACAGACTATCTGGAAGCAATGCACATGAT 4719
Db 1025 -----TyrSerGlySerAsnAspThrSer 1032
QY 4720 TACTTGAAACAGAAATTTACTGCTAGTTGTGTGATTAAGCTGGAAGTGCCTGTGTGA 4779
Db 1032 ----- 1032
QY 4780 CAAGAATATATACAAATTTGAGAAAAACAGCTTACTATGATGCCGACAAACATTTGTGG 4839
Db 1032 ----- 1032
QY 4840 TGCACAAAATTTATGAAAAATGACGACAAATATATCTAATCTTTCGAGTAAGTAAGTGC 4899
Db 1033 -----AspThrThrGlyLysGlnThrProSerSerAsnAspAsnLeu 1047
QY 4900 AAAAGATTAGTAAAGGCAACACAGGTCTATTAAATGTCGCAAAACAAAGGCTTAAT 4959
Db 1048 LysAsnIleValIleuGlnAlaSer----- 1055
QY 4960 AACTACAAATATCTTGAAGAATTTGACTGAAGATGTGCTTTTCTCTCTCGTCACTACGT 5019
Db 1055 ----- 1055
QY 5020 ATATGTTTTCATGATGATGCGAATTAATACAGATCCAGAAAGTTAAAGATGAAGATGGG 5079
Db 1055 ----- 1055
QY 5080 TTGCAAAAAGATTGATGAAGTGGCGCAACGGAAGGTACAATTTGGGTCAATACTAC 5139
Db 1056 -----GlySerThrGlu----- 1059
QY 5140 AAGAAAAAAGAAAAAGAAAAATTAATAAAGCTCGAGTCCGACAAATATCTTATGAG 5199
Db 1060 -----GlnLutLysGlnLysMetLys----- 1066
QY 5200 GTCCCGCCTGTGATGCTATGAATATAGTTTTATGATTTAAGATATATTTAGGT 5259
Db 1066 ----- 1066
QY 5260 ATTGATTAATTTGGAATGTAACAAAAACAAAGACGAGCAAAATTTGAGAAATATTAAAC 5319
Db 1067 -----GlnIleGlnAlaLysIleLysIleLeuAsn 1077
QY 5320 AAAAATGGAACATCAGTTGGCAAGAAGTATAGTACTACAGAAATCCCGTACTACT 5379
Db 1078 ---GlyAlaThrSer-----GlyValProValThrLysAsnSerValLysThr 1093
QY 5380 GCGCGAAAAATTTTCTGGAACGAAAAATTAAGAAATGTGTGTGGAACCATGATATCCGG 5439
Db 1094 ProGlnGlnThrPyrPyrGlnAsnIleAlaLysAspIleTyrPasnIleMetValCysAla 1113
QY 5440 -----TACAACGTGTAGGATGATGAAATATGTGAATATGTGCAAGAACTGATGAA 5493
Db 1114 LeuThrTyrLysGlu-----AsnAspAlaArg----- 1122
QY 5494 GATCTAAAAAATGTGGTCTGTACCTCAGATGATGATATCTATGGGAAAAATCCG 5553
Db 1122 ----- 1122
QY 5554 GATGAAGTACTGCGTATCATGTTCTTCGATGGTTGGCCGAATGGGGTAAAGATTTTGC 5613
Db 1123 -----GlyThrSer----- 1125
QY 5614 AAACATPAAGAAAAAGAAATTTGAGAAATTTGTAGGGCGGTGAATGATTTACTTGGT 5673
Db 1125 ----- 1125
QY 5674 GATTAATGAAGATTAAGAAAAAAGAAATGTACAGATCGGTACACAAATATAAAATTTAT 5733
Db 1125 ----- 1125
QY 5734 AGTGATGGAACACACAGTATGAAAAACAATCAAAAAATATGTGGAATTAAGCAAA 5793
Db 1126 -----AlaLysIleGlnIleAsnLysAsp--- 1133
QY 5794 ATATATTCGAGCATCCTGTGCAAAAGATGACAGAGACCTGCGCAATATTAGACAA 5853
Db 1133 ----- 1133
QY 5854 CAATTAAAAAATTTGTGAAAAATAAAGTGAAGATTGTAATATAAGTATGAAGAT 5913
Db 1134 ---LeuLysLys----- 1136
QY 5914 GTGTCCACACAGGATTAATGATGTAATATGTCMAAATATGCCCCGATCATTTAGACGAT 5973
Db 1137 -----AlaLeuThrPaspGlnAlaAsnLysAsnThrPro----- 1147
QY 5974 GAACCAAAAGAAAGTTGAAGAAAGTGAATTTGCAAGTGCACAGAGTCCACACAGTGA 6033
Db 1147 ----- 1147
QY 6034 CGAAGGAAACACCGTCACACAGGGTATCATGATATCAAAAGCAGCGCATGAAAAAA 6093
Db 1147 ----- 1147
QY 6094 GAAGCAAAAACAGCGCCGCTACAAAAACAGCCGAAAAAAGTGAATAATCTAACACAGAA 6153
Db 1147 ----- 1147
QY 6154 ATGCGACACAACACGAAACCGACGACGACACAACAAACAGAAAAACGAAATCAACA 6213
Db 1147 ----- 1147
```

Db	1283	-----LysCys			glnthr	128
OY	7288	TCACATCATCCGATGCTATGATATGCCAGAAATCGCTGATTAATGTCCT			GAA	734
Db	1287	glnbersnmsnsnslalabsnclupbeserhrythrlenclyalaserprothrAlala				130
OY	7342	GAATTTCACAAAGTGTGAGTGTCTCGAATCTTTCAAAAAAGGATCTATGATTCATACAAAA				740
Db	1307	gluphe-----LeuGlnLysleuGlySer				131
OY	7402	AAATTTACTAGACCTTAAATTAACCTATGAATTTGTGTAGAGAAAAGCAGATTAATTAATCT				746
Db	1315	-----CysLysAsnAspAsnGlyTyr				132
OY	7462	AAAGAACGAAATAATATATGATATTAACCTGAAGAAAATTTATATACCTATGAGTCT				752
Db	1322	-----GlnsncllyglunspAsnLyslleAspPheLys--AsnProAspLysThr				133
OY	7522	ACAAAGAAAAAGAAAGTAAAAATAGTTGAGCAATAATAATTAATCTTGAGATCTTCAAGAA				758
Db	1338	Phelysclunlahisler-----CysAspPro--Cys				134
OY	7582	CCTTATGACACCGTATTAATATATAGAGAAAGAAACCTTTGTGAAATAGAGAGAAAT				764
Db	1348	ProIleThrlGlyAllys-----CysGlnAsnGlnHis				135
OY	7642	CGTTTAAAGTAGATTAATGATGAAATGTTACAAAAATTCAAAGTTCTATCAGAGAAA				770
Db	1358	-----Lys				135
OY	7702	AAAGAGTATGCTGTACCTCCAGAAAGACACATATGCTTAAAGAAATTTAGATGAAT				776
Db	1359	-----CysValGlySerAlaAsnGlyLysGlyLysAsnAsn--LysIle				137
OY	7762	AAATTTGAAAGATTAAGATTAATTAATATCTCTAAAAATGGTGTGTGCAACTGCACAA				782
Db	1374	ThrAlaGlnAspIleLysAsn-----LysThrAspPro				138
OY	7822	AATGAAGAAATGACATATAATAAAAAATCTCACTCAGAGAACGGGTGCGCATGAAATCA				788
Db	1385	AsnGlyAsnIleGluMetValValSerAspSerThrAsn-----Lys				139
OY	7882	ATATGTGACTACTGAATAATATAGTTGTGGCTGTGTGATCACTAGTTAGAGAACAGAT				794
Db	1399	-----ThrPheGlnHisLeuGlyAsp-----Lys				140
OY	7942	ATGTTACGAATTTGTTGTTACTTACCTCCGTAAGAAATTAATTAATAGGTTTTGAA				800
Db	1406	-----CysLysSerSerGlyIlePheLys-----Lys				141
OY	8002	TTCATATATGAAAAATGAGAAATAAAAATTAAGGTAGAAATTAACAAGATGTACAA				806
Db	1414	GlyIle-----Lys				1415
OY	8062	ACGTTTCGTTCTGCTGTGGATGCTAATAGAAAGAAATTTTGCAAA--GCAATGAC				811
Db	1416	-----ArgLysAspGluTyrPheCysAlaAsnVal				1425
OY	8119	TGCAAGACACCAAGATGCAAAACTTTTATAGAAAAAGAAAGATGGATGATTTACACGC				817
Db	1426	Cys-----GlyAlaAspIle				1430
OY	8179	ATACATTAATACAAAGT--AAGTGTGACATTAAGAGCAGATCCACTGTTGATGATTAAT				8235
Db	1431	CysThrIleuGlnLysLyslleLysAsnGlyGlnGlnGlyAsp-----LysLysTyr				1447
OY	8236	ATACCTCAACGGTTTCGATGATGACTGAATGGTCTGGAATTAATTTATTAAGACACTAGT				8295
Db	1448	IleThrMetLys--GluLeuLeuLysArgTyrPheGlnLysIlePhe--Leu				1462
OY	8296	GAAATATGAAAAATTTAAAAATATCATGTATGATCACTGT--AAAACATTCAGACAGATGC				8352

DB 1463 GLuAsPtyrAsnArgIleAlaGlyLysIleLysLeuGlyThrLysLysGluAspGlyCys 1482
QY 8353 AAGAAATGATTTATGATGATAAATAAGTGAACAGTGTAAAAAGAGATGCAAGATATATAA 8412
DB 1483 Lys-----CysIleLysGlyCysIleGlu----- 1490
QY 8413 AATTTGTCTTAAATGGAATCTCTATTGATATACAAATAAATACAAAGAATTG 8472
DB 1491 -----LysTrp----- 1492
QY 8473 TATGAACAACCAATATATACAAAATCTCTACTTATGATCATGTGCAAAATTTGTACAA 8532
DB 1493 -----ValGln-----Glu 1495
QY 8533 AAGTGAATACTTTTAAAGTGAATGTCTGTGAGAGCTTTCTGAAATCTTCATGAA 8592
DB 1496 LysThrLysGluTrpGln----- 1501
QY 8593 ACAAGTAAAGTGTGATTTAATAATTAATGAAAATGAGTCTCTCCAAATATACGAACA 8652
DB 1502 -----LysIleAsnAsp-----Thr 1506
QY 8653 TATGCTTTGAGAAGAACACCAAAAAGTTATAAAGAAGCTTGACGTTGTACACTACTCT 8712
DB 1507 Tyr----- 1507
QY 8713 AAGAAATCATTTGATGATATTTGCTACCGATCAAAACAAGATGATGATGAATACAA 8772
DB 1508 -----LeuGlu 1509
QY 8773 ACTTTTACCTTCTGTCGAAGATGATGATTAATTAATCTGTATGATGGAACGATAC 8832
DB 1510 GlnTyr-----LysAsnAspAspGlyAsnThrLeuThrAsnHe----- 1522
QY 8833 CTGTCTTAAATAGTTCAGATGATTAACAAGGHTATGATTTCTCTCCAAAGAAGACAT 8892
DB 1522 ----- 1522
QY 8893 TTATGTACAGACCTATACATGATATATATATAGAAAAGATGATAAAGAAATTTTAAAA 8952
DB 1523 -----LeuGlnInPheGlnTyrArg----- 1529
QY 8953 AAAAACTTCTACTCTGCTTCTGCTTCACTCAAGCAACATTTGAGTCAAAAATATTAATCG 9012
DB 1529 ----- 1529
QY 9013 GAAGAAGATGTGCTTTGAGCAATGAATATAGTTATGACAGATTAATTCGATATATAT 9072
DB 1530 -----ThrluPheLysAsnAlaIle 1536
QY 9073 AAGAAGATGATATGATGACACTCATTTATCTGAAAAAATTAATAAATATTTTGAACA 9132
DB 1537 LysProCysAspGlyLeuAsp----- 1543
QY 9133 TCAATATGACAAACCAAAATCGTAAACATGGTGGAAAAATATATGACCTCAGATATGG 9192
DB 1543 ----- 1543
QY 9193 CACGCTATGTTATGTGATATATAAATTCCTACTTCAAAAGTAAACATATAGATGAAGATGG 9252
DB 1543 ----- 1543
QY 9253 TGTCAATTTACCAAGATGAGAAGAAATCACTTTCTGTTGTTAATTGAATGGGCA 9312
DB 1544 -----GlnPhe----- 1545
QY 9313 AAGCAAGATGTAAGAAAAACAATGTATGATTCATTAACAAACAAATGCTCTCGT 9372
DB 1546 -----LysThrSerCysGlyLeu 1551
QY 9373 TCAACGAAGATATATTTGAAGCGTCAGAAATTAATTAAGACAACCTGATGATGATGAT 9432
DB 1552 AsnSerThrAspAsn-----SerGlnAsnGlyAsnAsnAsnAsp 1564

QY 9433 ATTAGAAAAATATATAGCTTGAATATATGATTAATAAATAACATGGAATAATATAATA 9492
DB 1565 -----LeuValLeuGlyLeuLeuAsnLysLeuGlnLys 1575
QY 9493 AATATATGCAATTAATAAGATCAATCTTCAGTAAATATAGACAAATTAACCATCTGAGAA 9552
DB 1576 LysIleSerGluCysLysGlnGlnHisSerGlyGlnThrGlnThr-----ProCysAsp--- 1593
QY 9553 AATGTCAGCATATATTAATAAATCAAAAGATTCATGCGCTTTGAGCTTAAATGATATA 9612
DB 1594 -----AsnSerSerLeuSerGlyLysGluSerThrLeuValGlnLysPalaAspPylr 1611
QY 9613 AATGAATAGTTACAGAACCAAAAATTAATGAATAAT----- 9651
DB 1612 GlnGlu-----GlnAsnProGlnAsnLysValGlnInProLysPheCys 1626
QY 9652 ---GAATTCAAAGACTACTAAAAAATTAATATCTCGTTATATTTGTTGAAGATGA 9708
DB 1627 ProAspMetLysGluProLysLysGluAsnAspGluGluValGlyThrCysGlyGlyAsp 1646
QY 9709 ACACACAAAATCATGATGATGATGAAATATAT-----AAGAAGAAGACCAACAGTT 9762
DB 1647 GlnGluLysLysLysValGlnAspSerValIleGlnGlnLysGluGlnLysAlaIleAsn 1666
QY 9763 CGTCTTAAGCACTCTATTTCTTACACCCATGATGATTTCTTATCAAGCACCTTTA 9822
DB 1667 AlaProGlnLysLysProProLeuThrPro-----GluAlaProLys 1680
QY 9823 TTCTCAACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 9855
DB 1681 LysGluGlnLysValAlaProLysProProProProProLysLysArgIleLysThr 1700
QY 9856 AAAAAAT-----GATATATGAAGAAGATGATGATGATGATGATGATGATGATGATG 9900
DB 1701 ArgAsnValLeuAspHisProAlaValIleProAlaLeuSerSerThrIleMetTrp 1720
QY 9901 GCCTTAGGTTTG-----ATAGCGCTTCATTTCAATGAAAAAATTCAAATCG 9948
DB 1721 SerIleGlyIleLysPheAlaIlePheThrTyrrPheTyrrLeuLysLysLysThrLysSer 1740
QY 9949 TCTGTG---GACTTGTGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 10005
DB 1741 SerValGlyAsnLeuPheGlnIleLeuGlnIleProLysSerAspPylrAspIleProThr 1760
QY 10006 TTGGAATCCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10065
DB 1761 LeuLysSerSerAsnArgIleProTyrrAlaSerAspArgHisLysGlyLysTrpTyr 1780
QY 10066 ATATATATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10125
DB 1781 IleTyrMetGlnLysLysPheSerGlyAspGlu---LysTyrrAlaPheMetSerAspThr 1799
QY 10126 TCTGATATTTACTTCCCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 10185
DB 1800 ThrAspIleThrSerSerGlnSerGlnTyrrGlnGlnLeuAspIleAsnAspIleTyrrAl 1819
QY 10186 CCAGTAGTCTTAATATATAAATGATGATGATGATGATGATGATGATGATGATGATGAT 10245
DB 1820 ProGlySerProLysTrpLysTrpLysIleGlnValIleValLeuLysProSerLysAlaGsp 1839
QY 10246 ATACCAAGTAT-----GATACACCAAGTATGATGATGATGATGATGATGATGATG 10284
DB 1840 ThrGlnAsnAspIleHisAsnAspIleProSerAspIleProAsnSerAspThrProPro 1859
QY 10285 ACGAATATGATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 10344
DB 1860 Pro-----IleThrAspAspGlnTrpAsnGlnIleLysLysAspPheIleSerAsnMet 1877
QY 10345 TTA-----CCAATATCAGAACCAAAATTAATTAATTAATTAATTAATTAATTAAT 10395
DB 1878 LeuGlnAsnThrGlnAsnThrGlnProAsnIleLeuHisAspAsnValAsp-----Asn 1895

[illegible]

Db	294	CysArgCysAspLysAspLysGlyAlaAsnAlaGlyLysProLysAlaGlyAspLys	313
QY	919	-----GTTCCTACCAATTAGATATATGTCCTCAATATTACGTTGGTGCAGCA	969
Db	314	ValThrIleValProThrTyrPheAspTyrValProGlnTyrLeuAlaGlyPheGln	333
QY	970	TGGCGAGACAGCTTTGGCCCAAAAACAATATTAAATGAAACAGTCACGCTCGT	1029
Db	334	TrpIleGlnAspPheCysArgLysLysLysLysLeuGlnAsnLeuGlnLysCys	353
QY	1030	CGT---AATGACAAA-----GAACGCTTAATATTGATCATATGACATGATGTACG	1080
Db	354	ArgGlyLysAspLysSerAspGlnTyrArgTyrCysSerArgGlnGlyTyrAspGln	373
QY	1081	ACAACCTTTTGGAAAAAAGGATTTTGCATTGGATTAATTAAGTCATGACGCTTCGACT	1140
Db	374	GlnThrIleSerAlaGlyLysValArgMetGlyLysGlyCysThrAspCysPhe	393
QY	1141	AAATGCAAAAGTTTGAAGTTGGTTAGGACAACAAGAACGCTTTAAAAACAAAA	1200
Db	394	AlaCysGlySerTyrGlnAsnTrpIleAspAsnGlnArgLysGlnPheAspGlnLys	413
QY	1201	GAATAATATGAAAAAAGAAATACATATATTTATGCAACGATACAAATTTGTCAATAT	1260
Db	414	---LysTyrThrTyrGlnLysSerAspGlyGlyArgLysLysArgAlaGlyGly	432
QY	1261	ATTATAT-----AGTGAATATATTAACAATTTATGAAAACTTAAGCAACGCAATAT	1314
Db	433	ThrThrLysTyrGlnGlyTyrGlnLysSerPheTyrGlnLysLeuLysAsnAspLysTyr	452
QY	1315	GCAACTAATGACACTTTTAAATTTACTAATATGAAGAAGTATGTAA-----	1365
Db	453	GlyThrValAspAlaPheLeuGlnLeuLeuAsnGlnLysAlaCysAspIleThr	472
QY	1366	-----CGACA	1371
Db	473	AspLysLysIleAsnPheLysGlnValAsnSerGlyGlyValAlaGlyLysGly	492
QY	1372	TTACAGAGAGAAAAAGATATTACTTTTACTAACAGTGTGATGACAAAAGGATATTATAT	1433
Db	493	SerGlyGlyThrSerGlyAlaSerGlyThrAsnAsp---GlnAsnLysGlyThrPheTyr	511
QY	1432	CGTTCAAGATATTTGCCAAGTGTGCTCCGAGCTGGGGGCTAAATGTGATGTATTAATAAC	1491
Db	512	ArgSerGlnTyrCysGlnProCysProAspCysGlyValGln-----	525
QY	1492	ACACAAACATACATATGATCGTGAACGCGTAAAT-----AAT	1530
Db	526	---HisLysGlyLysAsnGlnTrpLysAlaGlyThrLysValLysLysMetArgTrpSer	544
QY	1531	GAACACTATTAACCTCCATGGGTGTGAACCTTACTAATATGACTGCTTTATAGTGT	1590
Db	545	LysLeuTyrTrpIleAsnGlnLys-MetValLeuLeuLysSerLeuLysValVal	564
QY	1591	AATGAACAGGTGATTTATACAAAAATTA-GAAATTTTGTG-----AACAGTC	1640
Db	564	LysAspMetMetIleLeuLysLysAsnTrpLysGlnPheCysLeuThrGlnAsnSerSe	584
QY	1641	A-----ACTAATTCAAAAGCTAAAAATATATCAAAA--	1671
Db	584	AspGlySerValGlySerValAlaThrThrGlyAlaSerGlyLysAsnSerGlnLysLys	604
QY	1672	-----TGGAATGCTATTTATTAAGATGAAGAAATATATATATAGATGAATACT	1715
Db	604	GlnLeuTyrAspGlnTrpTrpLysCys---TyrLysHisAsnGlnValGlnLysValAsnVa	623
QY	1716	GGACAAATAATCTGAATTCATATGAT-----	1743
Db	623	IleGlnGlyLysValGlnLysAspAspArgLysLeuLysGlyAlaGlyLysLeuCysIleLe	643
QY	1744	---AATGCTAAG-----	1752

Dd	643	uPAspnrProLysuLysAsnLysGluValSerGIuAlaLysSerGIAspAsnAlaLAs	663
Qy	1753	-----ATATATCATTCATTAATATTTTGTTCGAATATAGGCTACATATTTATTAAGGATAC	1806
Dd	663	pLeGlnLysThrPheHisAspPhePheTYTGTTPValAlaHisMetLeuLysAspSe	683
Qy	1809	TATTAAAGCGAAATGACAAA--CTTAACCTGTGTATTAATAT--ACAACCCAGCATTG	1866
Dd	683	rLLeHisrTrpArgThrLysAsrGlydeuLysSerCylIleSerAspGlyLysThiMelLysCy	703
Qy	1863	TATTGATCATGTATACAGAAATTCGTATGTATGTTCAGAGATGGGTTTAAACAAAAGAGA	1922
Dd	703	sArGAsnGlyCysAsnLysCysAspCysPheGluLysTrpValLysGlnLysGluLinh	723
Qy	1923	AGAAATGGAATAGTATAAGAAACCTGTCAACAAAAAAGAAATATACACATTCGTATTA	1988
Dd	723	rGluTrpLysProLleLysAspHisPheLysrThrGlnGlnGlyLleProGlnGlyTYTty	743
Qy	1983	TAGTAAATATTAATTAATCTTTTGAAGGTATTTTAAAGTTATGATTAACCTTGACAA	2044
Dd	743	I-----	743
Qy	2043	AGATGAGCAAAATGGAAGAAGACTTATGCAAAATATRAAAAAAGAAAAATGAGTTTC	2107
Dd	744	-----PheTh	745
Qy	2103	CAATTTGGAAATATATATAGGACATTTATAGAAATGCAATGACACTCTGTGTAGATCACT	2166
Dd	745	rThrLeuGlu-----LeuLleLeuLysLeuGlnPheLe	756
Qy	2163	AAAAAGAACTCCGACGATATGTATAAGACAATATACAAAAGCAAGCATGTGAAACATCCCA	2223
Dd	756	uLysGlu-----AspThrGlnGluAsnThrGluAsnSerLe	768
Qy	2223	TAAATGCAACAACAACCCGCTGTGTAAACCTCGTGGAGGACCCAAACCACTRAAAATAT	2288
Dd	768	uAspAla-----	770
Qy	2283	AAAAGAAATAGCACAAATACTTTTAAAGAGATGCATACAGAGACAGCAAAATCGTGTC	2344
Dd	771	-----GluGluAla-----	773
Qy	2343	TCATTAATTTGAAAGAAAAGGCACACGACGAAGTATATATAAACGTGGGGGTAGCAAAAAGA	2407
Dd	773	-----	773
Qy	2403	CTTCAGAGACAATTTATGTAGAAATATGATATAAACAATTTAACTGATATCTGTGTTTTTC	2466
Dd	773	-----	773
Qy	2463	AAATGGACCATGTGATGGCAAAAGGCACAGGTGATGTATACAAACAAGATTGTGCGTAGG	2522
Dd	773	-----	773
Qy	2523	AACTGAATGGGAAGTGGATCCGGAACAACATGCGTAAAGATCAGCAAGATGTATATATGCC	2588
Dd	773	-----	773
Qy	2583	TCCTAGAGAGACATATATGTATGATCATTCCAATTTGGACATTTCAACAACGGATGATCACC	2644
Dd	774	-----GluGluLeuLysHisLeuGln-----	780
Qy	2643	ACTTATATGTATATTTGTGTATGATTTAGTTAATTAATTCCTTTTGGGGAGTGTCTCT	2707
Dd	780	-----	780
Qy	2703	ATCAGCAAAATATGAGCAAAACAGATTAATACGATGTATTAAGAAAGATTAATACCTTAA	2766
Dd	781	-----LysLleLeuLysLeuGlnuAsnGlnu-----AsnAsnLeuAl	792
Qy	2763	GGGGCCCAAGAGATTAACCTGACCCCAAAACAACGACACATCTGTGTGAGCTATACGTTA	2822
Dd	792	aValValAsnAlaGlyThrGlu-----GlnLysThrLeu-----	803

QY 2823 CAGTTTGCAGANTATAGGTATATATTCGAGGAAGAGATCTCTGGGAAAGAAACGGTGA 2882
Db 803 ----- 803
QY 2883 CATGTAAAGCTGCAGACATTTTGAAGCTGTTTGGTAATATACATTAAGTCTCA 2942
Db 804 -MetaspLysleuasnHis----- 810
QY 2943 AGCGAAAGAAATGATTAATATATATGATGATGCCCCAAATATTTAAATTGAGGAAA 3002
Db 810 ----- 810
QY 3003 TTGGTGGAGCTATATAGACCAAGATGCGAAGCCAGAAATGATTAATAATATTT 3062
Db 811 -----GluLeuasn-----AspAlaThrLysCysLysAspCysProLe 823
QY 3063 G-----AAGGATAATCGGAGACACCAATCAACAACAAGATGATTTGGCGATATAGTGA 3116
Db 823 uProGluGluAspLysSerArgLysArgSer-----AlaAs 835
QY 3117 TCATACACCATTTGATGATTAATCCACAAAATTAAGATGATGACCGAATGGGCAGA 3176
Db 835 pProSerPro--AspIlePhePro----- 843
QY 3177 ATGCTACTGCAAGTGCAGAAAAGAGATATGATTAAGTTGAAGAGAGATGTAAGAGATG 3236
Db 844 -----ArgProGluGluLysGluAspAspGlu----- 852
QY 3237 TAAGATAAGAGATTAATGCTCAAGGCTGTACGAAGAGAGATGTAAGTTGTAAGAGT 3296
Db 852 ----- 852
QY 3297 CACAGAGCTGTAATGAATATATATATATATAGATTATGAAAGAAATGAATAT 3356
Db 853 -----AsnGluAspAspAsp----- 857
QY 3357 AATATCAGATAAATACAAAGAAATTAACAGACAGACAAATGCTGTTAGTAATAGTGC 3416
Db 858 -----GluAspGluValAlaArgAspAspGluGluThrAlaLysGluThrGluGlySerAl 876
QY 3417 TATGAGAGCTCCAGTACAGTCCAAAATTCATATAGACAGAGATGTAATGATTTTTC 3476
Db 876 ThrAspThrThrThrSerLeu----- 883
QY 3477 GGAATTTATCCAAACAAATGTGTGCGAAAAGTAAATTAAGTGTACTAGTGAAGTGC 3536
Db 884 -----AspValCysPr 887
QY 3537 TGTCAATGCTACTAACACACCGATGAAATGTGTGAGCATATCTCCATGATACAGGAA 3596
Db 887 olleValGlyLysValLeuThr----- 894
QY 3597 TTTTGAATGATTTGTCACTGACAAATAGTTTGTGATGAAAAAGTGATGTAAGATTA 3656
Db 895 -----LysAspAs 897
QY 3657 CGAATAATATATGCTTTAGAGATAAACACAGACCATGATGCTGCTGTTGTAAG 3716
Db 897 nGluSerLeu-----GlnAsp-----AlaCysSerLeuLysTy 908
QY 3717 TGGATCGAAGCGACAGAGGTACAGATTAACGAAAAAAGCGGAGAAAGAGATAC 3776
Db 908 rGlyLysAsnAsnSerArgLeuGlyTyrArgCysValThrProSerGlyGluProThr 928
QY 3777 GGAATGTAACACGTGAATGATATATTAAGAAAAAGATGAAAAAGAAACAGTAGAAGA 3836
Db 928 r----- 928
QY 3837 TTGTATCCAAAAAAGATATATGATATCCGATTTGCCAATGGCAATATTAATTT 3896
Db 929 -----SerSerAspLysAsnGly----- 934

QY 3897 AGTGAAGACCTCGTGTGTATGCCCCCTAGAACAAAGTATAGCTACTTCT 3956
Db 935 -----AlaIleCysValProProAlaArgAlaGluLeuTyrIleLysIle 950
QY 3957 G-----CCAAATGATTAATTAAGAAAGCTTCATCAATC 3983
Db 950 eValAspTrpAlaThrLysThrGluSerProGlnAlaSerGlySerGluAlaSerThr 970
QY 3984 A-----CAATCACAAGTAAATTAAGAAAGCTTCATCAATC 4022
Db 970 rSerGlySerThrThrProProAspSerLysGluAlaLeuLeuLysAlaPheValGlu 990
QY 4023 TGCAGCAGCAGAAACATTTCTCATGTGATATATTA 4064
Db 990 rAlaIleArgLysThrPhePheLeuThrPheIleArgTyrLysGluGluLysAlaValAl 1010
QY 4065 TAAAGATGTGAGAAATGAATC----- 4089
Db 1010 eGlnGluGlyAlaGlyHisGlyLeuProArgValGluGluLysSerProGluTyrAspPr 1030
QY 4090 -GATAAGATTAAGAAAGCAAAATCCTCCGCAATTTTGAGATCATGTTCTACAC 4148
Db 1030 oGluAspLysLeuLysGluGlyLysIleProAspLysLeuLysArgGlnMetPheTyrTh 1050
QY 4149 ATTTGAGATTAATAGATTTTATTTTGAACAGATATATCAAAAAGGTCATGCTGAGGG 4208
Db 1050 rLeuGlyAspTyrArgAspIleLeuPhe----- 1059
QY 4209 AAGTAACTTAAAGCAAAATAGATCTCTTTCAAAAATGCTGACCAAAAATCTCCTAA 4268
Db 1059 ----- 1059
QY 4269 TGGAAAAACAGCCAAAGATGTGGACAGAACATAGCATGATATGGAAGCTATGCT 4328
Db 1059 ----- 1059
QY 4329 ATGTGCTACTATAAATTTGGGGCAAAAAGATGATTTACGAAAACATACGGTTACAA 4388
Db 1059 ----- 1059
QY 4389 CAACTCAATTTAGTGCAAAAAGACACACCTTTGGAGAAATTTCCAAAGACCCAGTT 4448
Db 1059 ----- 1059
QY 4449 TTTACGATGCTAACCGAATGTGTACAGACTATTTGCTATACAGCAAAAATTTTGA 4508
Db 1059 ----- 1059
QY 4509 CGATGTGACGAAAAATGTAAATGCAAAATGCAAAATGCAAGTGTCAATGTAATA 4568
Db 1059 ----- 1059
QY 4569 GAAATGCGAGACTACGTTAATATATGAAAAAAGAGTGCATTCACAAAGTAA 4628
Db 1059 ----- 1059
QY 4629 ATATTAAGAGATGACGCAACAAAAAAGATTCGATAGACAACACATGTGTAAATGT 4688
Db 1059 ----- 1059
QY 4689 TACAGACTATACGTGAACGATGCAACAGATTAATCTGGAACAGAAATTTACTGCTAGT 4748
Db 1060 -----SerLysSerAsnAspThr-----ThrSerValSe 1069
QY 4749 TGTGATACCTGGAAGTCTCTGTGTGTACAAAGAAATATATCAATTTGTATAAAAACA 4808
Db 1069 rLysAspThrProSerSerSerAsn----- 1077
QY 4809 GCGTTACTATGATCCGCAACAAATGTTGGTGCACAAAATTTATGAAATAGACGACA 4868
Db 1077 ----- 1077
QY 4869 ATATACTAATTTCCAGTAAAGATTAAGTGCAAAAGATTAGTAAAGAGCAACACAGG 4928

Db 1077 ----- 1077
QY 4929 TGGTATTAAGTGGCAAAACAAAGGCTCTAATACTAATACTGTAAGAAATTGACTGA 4988
Db 1078 ----- 1084
QY 4989 AGATGCTCTTTTCCTTCCTGCTGACTACGTATATGTTTCATGCATTTGGATGGCAATTA 5048
Db 1084 ----- 1084
QY 5049 TACAGATCCAGAAATTAAAGATGAAATGGCTGCCAAAAAGATTGATGCAAGTGGCGGC 5108
Db 1085 ----- 1090
QY 5109 AACGGAAGGTTACAAATTTGGGTCAATCTACAAAGAAAAAGAAAAAGAAAAATTA 5168
Db 1090 rthrglu ----- 1098
QY 5169 AACGTGGATGCCACAAATATTTCTATAGAGTCCGCCCTTGTATGTGAATATATAG 5228
1099 ----- 1101
QY 5229 TTTTATGATTTAAGATATATATCTAGTATTTGATATTTGGAAGATGAAAAACAAA 5288
Db 1101 ----- 1101
QY 5289 GACCGAGAAAAATTTGAGAAAAATTTTACAAAAATGAAACATCAGTTGGCAAGGAG 5348
Db 1102 ----- 1116
QY 5349 TGATAGTACTACAGAAATCCCGTAGTACTGGCGCAAAATTTTCTGAAACCAAAATA 5408
Db 1116 aProAsnLeuValSerHisPro ----- GluThrTrpIleuSnsngl 1131
QY 5409 GGAATGTGTGTGAACCGCATGATAGCGGTACAAACGTTGATGGATGGAATAG 5468
Db 1131 yLysTrpIleThrHisGluValCysAlaLeu ----- 1142
QY 5469 TGGAAATAGTCAAGAAAGTATGATGAAATCTTAAAAATGTGTTCTGTACCTTCAGATGA 5528
Db 1142 ----- 1142
QY 5529 TGATTAATCTATGGGAAAAAATCCGATGAAGGTACTGCCATCAGTTCTTCGATGGTT 5588
Db 1142 ----- 1142
QY 5589 TGGCGAATGGGTGAAGATTTTTCACAAACATTAAGAAAAAGAAATTTGAGAAATTTGATAG 5648
1142 ----- 1142
QY 5649 GCGGTATATGATTACTTGTGTATATGAAGATTAAGAAAGAAATGTACAGATGC 5708
Db 1142 ----- 1142
QY 5709 GTGTACACATTAATAAATTTATAGTGTAGTGAACACACAGTATGAAAAACAATCA 5768
Db 1142 ----- 1142
QY 5769 AAAATATGTAATTAAGACAAATATTTTCGAGCATCTGTGGCAAAAGATGCAGA 5828
Db 1143 ----- 1153
QY 5829 GAGAGTCCGCAATATTATAGACAAACAAATTAATAAATTTGTGAAATTAAGATGGAGA 5888
1153 u ----- 1153
QY 5889 TTGTGAATTAATGATGTAAGAGATGTCCACACAGCGATTAACGTATGTAATAGTCA 5948
Db 1153 ----- 1153
QY 5949 AAATATGCCCGCATCATTAAGAGATGAACCAAAAGAAAGTTGAAGAAAGTGTAAATTGTCA 6008

Db 1153 ----- 1153
QY 6009 AGTGCCACGAGTCCACACAGGTGTACGAAGGAACACCGTCAACAGGGTATCAATGAT 6068
Db 1153 ----- 1153
QY 6069 ATCAAAAGCGAGCATTCGAAAAAGAGGAAAAAGAGCCGCCCTACAAACAGCCGAA 6128
Db 1154 ----- 1157
QY 6129 AAAAGTGGAAAAATCTAACAAACAGAAATGCGAGCAAAACAGAACCCGAGCAGACACA 6188
Db 1157 nLysIleGluAsn ----- 1161
QY 6189 ACAAAACAGAAAAAGAAATCAACAGACACACACAGAAATGTAGGTGGCAATGAT 6248
Db 1161 ----- 1161
QY 6249 AAAGCCATTTCTTTCGATTAACCAATACAGATACAGGGGTGAAGGTTTATCCAA 6308
Db 1162 ----- 1174
QY 6309 AACGTATGACAAATATCTTAATGGGTTGTATTTAGTAACTTAAGAAAAATGAAA 6368
Db 1174 s ----- 1174
QY 6369 TGGCATATGTATGCTCTCTCTAGSAGAAAAAATTTATATATATATATATATATTA 6428
Db 1175 ----- 1182
QY 6429 TTATGAACCTGAATAATTAAGCGTACAAATGATATAAAGAGCTTTTATTAATGTCCAG 6488
Db 1182 nValLysLeuAsp ----- 1186
QY 6489 AATAGAACTCAATTTTGTGTTAAATATATATATTAATTAATTAATTAATTAATTA 6548
Db 1187 ----- 1188
QY 6549 ATTGAAATGGAACAATTCAGATGATTAATAAGAAATATATATATATATATATATATTA 6608
Db 1189 ----- 1197
QY 6609 TTATTAAGATATGTTTTTGGAACTGATATTTTAAATGATTAATAAATTAATTAATTA 6668
Db 1198 ----- 1203
QY 6669 AAATAGTACAAACCATTTCTCAATGAAATTAATAAGAAAAACAGATTAATAAAGAA 6728
1204 ----- 1208
QY 6729 TGAAGAATTACGTAAATATTTTGGAGAAAAATTAATAAATTTATTTGGAGAAATGAT 6788
Db 1208 ----- 1208
QY 6789 ATATGATTAATCTTATCATCTACAGACGAAAAAGAAAAAATTAAGATATTA 6848
Db 1208 ----- 1208
QY 6849 CCAATGACATGACATGACCAAACTGACGCTTCCCTTGAAGAATTTGTAATAAGGCCCA 6908
1209 ----- 1217
QY 6909 ATTTTGAATGGTTCACAGAAATGGGAGAAATTTTGAATTAAGAGGAAGAAAGAT 6968
Db 1217 rTyrrPheArgTrpPheGluIuTrpGluSerPheCysArgGluArgLysLysArgLys 1237
QY 6969 GTTAAATTTGGAGCGGCTGTAAAGAAATATGATGATGATGATGATGATGATGATGATGAT 7013
1237 ----- 1255
QY 7014 TAATGACGTTAAAGACACAGAA ----- 7035
1255 rGlyAspGlyGluAlaCysAspSerIleSerThrHisAspTyrSerThrValProSerPh 1275

QY 7036 -----TGTGCAAGGCGTGTGTAACATATCAAAATTTTATTAAGAAGTGAA 7082
Db 1275 easnCySPROGlyCysglYshIScysSerSerYArGLytrPllEGluRgYsLsly 1295
QY 7083 AACGATATTAAGAAAGCAAGAAAGAAAGTCAAAAGAGATTAAGTGGCAAAAGTATA 7142
Db 1295 sllEGluPhelHleYsgInserNslnAlArYGLYgIn ----- 1307
QY 7143 GGATTAATCTTCTGAGAAAGACATAGAGCAACATGTCATCATGATATTATAA 7202
Db 1308 -----GlnYsThrAspAlaThrArgAsnAsnGlyAs 1318
QY 7203 CATGAATTAAGAAATATATGTGCAATTAAGATTTGTTCTTGATGCAAAAACCTTCTC 7262
Db 1318 nThrPheAspLysgluPhelHleYsgIn ----- 1325
QY 7263 ACACTACCAAAAAACAACAACATCAACATCCGATGCTAATGATATGCGCAANTC 7322
Db 1326 -----LysThr 1327
QY 7323 GCTGATATTGTTCTGAGAAATTTAACAAGTGTAGTGTCTGAACTTTCAAAAAAGG 7382
Db 1327 rleuGluThrTrpProAsp ----- 1333
QY 7383 ATCTATGATTCATACAAAAAATTAAGTACACTAATACATGATTTGTAGAGA 7442
Db 1333 ----- 1333
QY 7443 AGCAGATATTATTATCTTAAGAACAGAAATAATATGATATTACCTGAGAGAAA 7502
Db 1334 -AlaAlaLysPheLeuGluArg ----- 1340
QY 7503 ATTTATACCTATTGACTCTACAAAGAAAGAAAGTAAATAGTTGAGCTAATATAA 7562
Db 1341 -----LeuLysAsnGlu 1344
QY 7563 TCCATGCGATCTTAAGAAACCTTATGACCTGATTAATATATAGAGAAAGAACCTTG 7622
Db 1344 yProCysLysThrAsnLysgluTrpGlyLys ----- 1355
QY 7623 TCAAAATAGAGAAAGAAATCGTTTAAGGTATATGAAATGTAATACAAAATTC 7682
Db 1356 -----AspIleAspPheGlu -----LysAspSe 1363
QY 7683 AAAGTTCTATGAGAGAAAGAAAGATATGTACTCT-----CCAAGAGAGAAATATGTG 7739
Db 1363 rlyshThrPheGlnHlsThrGlu--TyrCysglYProCysProLysPheLysThrAsnCy 1382
QY 7740 CTTAAGATTTAGATGAATTAATTAAGTAAAGTAAAGTAAATTAATCTCTAA 7799
Db 1382 s ----- 1382
QY 7800 AATGTTGTCGACTGACACCAATTAAGAGATAGACATATATAAACTTCACACTCA 7859
Db 1383 -----Glu 1383
QY 7860 GAACGGG-----TGCGCAATGATCCATATATGTAT 7890
Db 1383 nasnCyLysnCySglYValSerGlyLeuAsnGlyAsnCySAspGlyAspLysSerIleAs 1403
QY 7861 -----ACTATGAATATATAGTTTCGCTGATCTGGGTGACATAGTAGAG 7934
Db 1403 PALALysgluIleAlaLysMetArgSerSerThrThrAspAla-----ValMetArgVa 1421
QY 7935 AACGATATGTTAGCAATTTGGTGGTACTTACCTCCCTAGAAATTAATATATATAGT 7994
Db 1421 lSerAsp----- 1423
QY 7995 TTTGAATACATATATGAGAAATGAGAAATATAAAGTGAATATAATACACGA 8054
Db 1424 -----AsnAs 1425

QY 8055 TGTACAAAGCTT-----CGTTCTGCTGTGGATGCTAAT----- 8091
Db 1425 pThrAsnThrPheGluGlyAspAspLeuLysAspAlaCysGlnHlsAlaAsnIlePheLys 1445
QY 8092 -----AAGAAAGATATTTGGAAGCAATGACGTCAAGCAACCAAGATGCAAAACT 8144
Db 1445 sglYIleArgLysAspValTrpLysCys----- 1454
QY 8145 TTTTGAAGAAAGAGAGATGATGATTTGAA-----CGCATACATTAATATCAAGATTA 8198
Db 1455 -----GlyTyrValCysglYValAspIleCysgluGlnThrAsnIleAsnGluArg 1471
QY 8199 GTGTGACATTAAGAGCAATCCACCTGTGATGATATATATACCTCAACGGTTT----- 8250
Db 1471 gThrAspGlyLys-----GluYrIleGlnIleAlaGluLeuPheLys 1485
QY 8251 -CGATGATGACTGAAATGCTCTGAATATTATTAAGACACTGATGAGAAATTTGAAA 8309
Db 1485 sarGTrpValGluAsnPhe-----LeuGluAspTyrAsnLys 1497
QY 8310 ATTTAAAAATCATGTGATCACTGT-----AAAACATCTGACAGATGCAAGATTA 8360
Db 1497 sIleAsnAspLysIleSerHlsCysIleLysLysglYgluLysSerLysCysIleAsnGlu 1517
QY 8361 TTATGATGAATAATAGTGTGACACTGTAAACGATGTCAAGAAATTAATAATTTTGT 8420
Db 1517 yCysgluLysAsn-----SerLysCysLeuGluLysTrpIleGluLysLys----- 1533
QY 8421 TCTTAATGGAATTCCTATTCGATATACATCAATTAATTAACAAAGATTTGTAGAA-- 8478
Db 1533 eAlaGluTrpGluAsnIle-----LysLysArgPheAsnAspGlnTyrGluAs 1549
QY 8479 -----CAACCAATATATTAACAAAATCTCTACTATGATCATGTCAAAATTTTGTACA 8531
Db 1549 nLysAspGlnProAspTyrAsnValLysSerIle-----LeuGluGluLeuIlePr 1566
QY 8532 AAAGTTGAAAACCTTTTAAAGATGAATGTTCTGTGAGAGCTTTCTGAAATATCTTCATGA 8591
Db 1566 oLysIleAlaValAlaAsnAspGlnAspAsnValIleLysLeuGlyValPhe-----Glu 1584
QY 8592 AACAGTATGCTTTGAATTTAATTAATTAAGAAATGATGCTCTTCATATACCAAC 8651
Db 1584 uAsnSerLysglYsThrLeuIleSerAsnThrGlnAsnAsnLysgluAsn----- 1601
QY 8652 ATATGCTTTGCAAGAAACACCAAAAAGTATTAAGAACCTTGACACTACCTTC 8711
Db 1602 -----AspAlaIleAspCysPheLeuLysLys 1610
QY 8712 TAAAGAT--CATTTGATATATGT-----CTTACCGATCAAAAAGAGTGATG 8759
Db 1610 sleuGlyValLysAlaLysAsnCySPROGlyLysProSerGlygluLysgluInserAspCy 1630
QY 8760 TAAAGATTTCAAACTTTTACCTCTGCTCGAAGAAAGATTAATGATTAATCTTGATA 8819
Db 1630 sLysgluProProleu-----ProAspGluGluAspGlnAsn----- 1643
QY 8820 TTGAGACGATACCTGTCTTCTTAATAGTTCAGATGATTAACAAGGTGATATCTCTCC 8879
Db 1644 -----ProGluGluAsnThrLeuGluProPr 1652
QY 8880 AAGAAAGAACATTTATGTAACAGACCTATCACTGCATATATTAAGAAAGGTGATA 8939
Db 1652 oLys-----PheLysPro--ProThrThrGlnProProGluGluLysGly----- 1666
QY 8940 AGAAATTTAAAAAAAACCTTCTTACTCTGCTTCAGCAAGCAACATTTGATGTC 8999
Db 1667 -----GlyGluThrCysglYAs 1672
QY 9000 AAAATATAATTCGAGAACAGTTGCTTTGAGGCAATGAATATATATATACAGATTA 9059
Db 1672 nLysgluGluLysLysAspLys----- 1679
QY 9060 TTCGATATATTAAGAAACATATGATGACACTTCATTTCTGAAAAATTAATAA 9119

[illegible]

SO	Sequence	1726	AA;	
Alignment Scores:				
Pred.	No.:	4,916-79	Length:	1726
Score:		1244.50	Matches:	501
Percent Similarity:		34.5%	Conservative:	276
Best Local Similarity:		22.258	Mismatches:	632
Query Match:		6,408	Indels:	843
DB:		17	Gaps:	88

[illegible]

```

259 CCAATGATTAATTAAGTTCATTAAGGAACAACACAGTAATTAATTCAGATGATGATGATGGAATTTGAGA 318
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 61 ProCys1LysIleu---TyrGIuAsPTrTyrTAsnAsnIArgIeuls-----ArgLysArg 77

QY 319 CACCTCTTCGCATGGTAGGAGAACAAACCCAGTTTCATAGATGAGACAAATCTGCATGTGGGA 378
      ::||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78 TyrProCysAlaAsnArgGIuInThIValArgPheSerAspGIuTyrGIdyGIuIcysThr 97

QY 379 ----ATAAAATACGTAATTTATAAAGAAAAAATGATGCTATA--GCCTGTGGCGCACT 432
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 98 PheAsnArgIleLysAspSerGIuAsnAsnAspAsnSerIeuGIuAlaCysAlaProTyr 117

QY 433 AGAAGACGACATATGTGTGATTAATAAAGCTTGGAAGCTCTAATATATTAATTAACCCAAAT 492
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 118 ArgArgIeuIleuLysAspTyrAsnIeuGIuLysMetGIuLysThrSerThrIleLys 137

QY 493 ATTCATGATTAATTTGGGAATGTACAGTACAGCAAAATACGAAGGTGGAATGTAATTTGTT 552
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 138 ---HisGIuLeuIeuAspValCysMetAlaIalLysTyrGIuIAspSerIleLys 156

QY 553 AATATATCT---CCACATAAGAGACT-----TCAGACGGCTTACTAGCT 594
      :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 157 ThrHisTyrThrIuHisGIuLeuIleuThrAsnProAspThrIuLysSerGIuIeuCysThrIle 176

```

QY	595	CTTGCAAGAAAGTTTGGCAATCTAGCTGATTTGTAACAGAAATGATATCTTT-----	648
Db	177	LeuAlaArgSerPheAlaAspIleLysPleValArgGlyLysAspLeuTyrLeuGly	196
QY	649	-----AAACAAATGTCATGACAAAGATAGAAAGGGCTCCGAGAGGTTTC	696
Db	197	TyrAspAspLysGluLysAspLeuIarGlyLysLeuGlnAsnLeuIleGluIlePhe	216
QY	697	AAGAAATATCATGATGAGATGGAGATGAGATTAATAAATGATTTCAATCCTGATGATCT	756
Db	217	LysLysIleIleHisGluAsnLeuGlyThrGlnAspAlaLysAspHisTyrLysAspGlu	236
QY	757	GGAATATTATTAATTAATTAAGAAAGAACAGTCGATGATGCAATGAATTAAGTATGGAA	816
Db	237	GluAsnTyrTyrGluLeuArgGluAspPrrPrrPrrHisAlaAsnArgSerThrValTrrPlys	256
QY	817	GGTATACAACTGGATGCATCATATTAATCTGGATATTTTATGCATACAGAAATATACA	876
Db	257	AlaIleThrCysHisAlaGlyGluSerAspLysTyrPheArgTyrThrCysCysSerGly	276
QY	877	CCATTAATTTTCAAAATCTCT--AAATGGCGCATTAACAGAGAAAG-----	918
Db	277	GluTrrPrrHisAspAspLysCysArgCysLysAspLeuIuGlyLysAsnGluTrrHisnGlu	296
QY	919	GTTCTTACCATTTTAAGATTATGTCCTCCATATTTTACGTTGTTTCAGCAANTGGGAGAA	978
Db	297	ValPrrThrTyrPheAspTyrValPrrGluTyrLeuArgTrrPrrHeGluGluTrrPrrAlaGlu	316
QY	979	GAGTTTCCGCAAAAAGAAATATTAAATGTAAGAAAGTCAAGACATCTCTGTATATGAC	1033
Db	317	AspPrrCysAlaGlyLysAlaGlyLysLysIleLeuLysnAlaIleLysAsnCysAlaGlyGly	336
QY	1039	AAA-----GAACGCTTATATTGTAGCATATATGACATATGACATGATTGTACGACATATTGG	1099
Db	337	LysGluAsnGluIarG--TyrCysAspLeuAsnGlyLysAsnCysGluGluTrrHisAlaArg	355
QY	1093	AAAAAGATTTTGGCATTTTGGATTAATTAAGTGTACGTGACGTTCGACTAAATGCAAAATT	1153
Db	356	GlyAlaGluIlePheValLysGlyAspAspCysHisLysCysSerValAlaCysAspArg	375
QY	1153	TTTGAAGTTGGTTAGGAGATCAACAAGACATTTTAAACAAACAAACAAAGAAATTTGAA	1213
Db	376	PheValLysTrrPrrLeaAspAsnGluArgGlyGluPheAspLysGluLysLysTyrAsp	395
QY	1213	AAAGAAATACAA-----TCATTTTATCGACGATACAAATTTGTCAATAAT	1266
Db	396	GluGluIleAsnLysThrHisGlyTrrThrIleThrThrGluAsnGlyLysIleAsnAsn	415
QY	1261	ATTATATAGTGAATTTTAAACAATTTTATGAAAACTTAAGGAAACGCAATATGCAACT	1322
Db	416	Leu-----TyrValGlyHisPheTyrLysIleLeuLysLys---TyrTyrPrrThr	431
QY	1321	AATGACACTTTTAAATTTACTTAATTAAGAAAGATATGTA---GGAGATTA	1374
Db	432	ValAspLysSerLeuGlnLysLeuAsnAspGluAlaIleCysLysLysPrrHisnAl	451
QY	1375	CCAGAGAGAAAGATATTACTTTTACTACAGTCGATGACACAAAGCATATTTTATGCT	1433
Db	452	GlyAsnGluLysAlaSerThrValAspPheAsnAsnGluValAsnThrThrPheSerHis	471
QY	1435	TCAGAAATATTTCCAAAGTGTCCCGACGTCGGGGTCAAAATGTCAT-----GGT	1488
Db	472	ThrThrTrrCysGluIarCysPrrThrPrrCysGlyAlaGlnLysGluLysAsnGlyLysGly	491
QY	1483	ATTAATAATCACACAAATTA---GATTAATGATCGTGAACGTGTAATATGACATAT	1539
Db	492	TrrPrrLysAlaLysGluLysSerCysAlaLysLysLysGluIarGluPheAsnLysGlu---	510
QY	1540	AAACCTCATGGGTGGAAGCCTATATATATACATGCTCTTATTAATGCTATGACAA	1599
Db	511	-----AsnSerThrAspIleLysIleLeuThrPrrGluLysGlyArg	524
QY	1600	GGTGATATTACACAAAAATTTAGAAATTTTGTATACAGCTCAACATTAATTAAGATATA	1659

```
Db 525 SerLysThrLeuGluLysLeuLysThrPheCysLysAspGlyGlnLysIleLys----- 542
QY 1660 AATTAACCAAAAATGGAGTCTATTATAGATGAATAATATAATAGTAACTGAA 1719
Db 543 ---AspAspIleThrLysCysHisIleAspAspGly-----ThrAsp 556
QY 1720 CAAATACAGCAATCAATATGAT-----AAT 1746
Db 557 AspGlnThrAspAspSerAspGlnValLeuGlyAspTyrPheLysLeuThrLysGlu 576
QY 1747 CCTAATATATATCATTTCAATATTTTGAATATATGGTTACATATTTATTAAGGAT 1806
Db 577 AspLysIleMetSerLysAsnAlaPhePheTyrPheValHisAspMetLeuIleAsp 596
QY 1807 ACTATTAGGCAATGCAAACTTAACCTGTATTAAT---AATCAACCCAGCTTGT 1863
Db 597 SerIleLysTyrPheAspGlnHisGlyArgCysIleAsnLysAspLysGlyThrCys 616
QY 1864 ATTGATGAATGTACAGAAATTCCTTATGTGTTCAGACATGGCTTAACAAAAAGAA 1923
Db 617 IleLysGlyCysAsnLysCysIleCysPheGlnLysTyrValGlnLysLysThr 636
QY 1924 GAATGAATAGTATAAGAACTGTCACAAAAAAGAAATATACAGCATGTATTAAT 1983
Db 637 GluTyrGlyLysIleLysAspHisPheArgLysGlnLysAspIleProLysAspTyr-- 655
QY 1984 AGTAATATTAATATCTTTTGAAGTTATTTTAAAGTATGATTAACCTGACAA 2043
Db 655 ----- 655
QY 2044 GATGAACAAATGGAAGAACTTATGAAATATTAATAAAGAAAAAATGAGTTTCC 2103
Db 655 ----- 655
QY 2104 AATTGGAAATATATAGCGATTTAGAGAAATGCAATA-----GACTCTTGTAGAT 2157
Db 656 -----ThrHisAspAspPheLeuGlnThrLeuMetLysAspLeuLeuGln 672
QY 2158 CACTTAAGAAATCGCCAGATATGTAAAGACATATATCAAAAGCATGTGAAGA 2217
Db 673 IleIleGlnAspThrTyr-----GlyAspAlaAsnGlu----- 683
QY 2218 TCCCATATATGCAACAAACCCGTGTGTTAAACCTGTGAGGACGCAACCCACTAA 2277
Db 683 ----- 683
QY 2278 AATTAATAAGAAATATACACATCTTTAAAGAGATGCATACAGAGACCAATCGT 2337
Db 684 ---IleLysArgIle-----GluAlaLeuLeuGlnAla----- 694
QY 2338 GGTCTTCATTAATTAAGAAAGGACACAGAGATATATATAAGGTGGGGTAGAGA 2397
Db 695 GlyValGlyClyIleAspPheAlaIleAlaLeuAlaGlyLeuTyrThrLysGlyPheValAla 714
QY 2398 AAGCACTTCAAGACAAATTTATGTAGAAATATGATTAACATTTCTAATCGTAACTTGGT 2457
Db 715 Glu-----LysAspThrThrIleAspLysLeuLeuGlnHisGlnLys----- 729
QY 2458 TTTTCAATGACCATGTGATGCGCAAGGACAGGTGATGCTATACAAACAATTTGTC 2517
Db 729 ----- 729
QY 2518 GTAGAACTGATGAGAGTGCATCCGGAACACATCGTTAAGATCAGGAAGTGTAT 2577
Db 730 -----GluAlaAspLysCysLeuLysThrHisAspAspThr 742
QY 2578 ATGCCTCTCAGAAAGACACATATATGTACATCCAAATTTGGAACATTTACAAAGCATGAT 2637
Db 743 CysProProGln-----GluAsp 748
QY 2638 CACCACTTAATGTATATATTTGATGATTTAGTTAATATTCCTTTTGGGGGATGTT 2697
Db ----- 2697

Db 749 ArgSerVal----- 751
QY 2698 CTTCATACGCAAAATATGAGCAACAGATATATACGATGATATAAGAAAAAGATATAC 2757
Db 752 -----AlaArgSerGlnSerAlaThrVal----- 759
QY 2758 CTAAAGGCCCCAAAGAGTAACTACGCCCCAAACACAGACAACATATCTGACCTATA 2817
Db 760 -----ProSerProAlaAspProLys----- 767
QY 2818 CGTTACAGTTTTCAGATATAGGTGATATTAATTCGAGGAAGAGATCTGGGAAAGAAC 2877
Db 767 ----- 767
QY 2878 GGTGACATGGTAAAGCTGCAAGACATTTGAAACGTGTTTGGTATATATACATAAGTCA 2937
Db 767 ----- 767
QY 2938 CTCAAAGCAAGCAAAATGATTAATATATATGATGATCCCCCAATATTTAAATTTGAG 2997
Db 767 ----- 767
QY 2998 GAAATATGGTGGAACTATAGAGCCAAAGTATGGGAAGCCATGAATGTATATAA 3057
Db 767 ----- 767
QY 3058 TATTTGAAGGATTAATCGGGACACCAATCAACAAAGTAGTTATTCGGATATAGTAT 3117
Db 767 ----- 767
QY 3118 CATACACATTTGGATATATATATATCCACAAATTAAGATGATGACGAAATGGCAGAA 3177
Db 767 ----- 767
QY 3178 TGCTACTGCAAGTGCAGAAAAAGAGTATGATTAAGTGAAGAGAAAGTGAAGGTCT 3237
Db 767 ----- 767
QY 3238 AAGGATTAAGGATTAATGCTCAGAGCTGTACAAAGAGAGTGTACAGGTGTACGAAGTGC 3297
Db 767 ----- 767
QY 3298 ACAGAAAGCTTTGATTAATGATTAATATATATATATAGATTATGCAAAAGCAATGGAATATA 3357
Db 767 ----- 767
QY 3358 ATATCAGATTAATAACAAAGATTAATACATGAACAAAGCAAAATGCTGTATATAGTGT 3417
Db 767 ----- 767
QY 3418 ATTGAAGCTTCAGTACGCAAAATATCATATATGACAGAAATGTATTAATTTTGTGC 3477
Db 768 -----AlaThrGlnGlnValAlaAspAlaAsn----- 775
QY 3478 GAATTAATCCAAACAAATAGGTGCAAAAGTAAATAAAGTGTAGTATGATGAAGTGTCT 3537
Db 776 -----AlaSerSerAspGln----- 781
QY 3538 GTCATTTGTACTTAACACCAAGTATGAATGTTGGAGCATATCTCATGATACAGAAAT 3597
Db 781 ----- 781
QY 3598 TTTGATGATTTGTCAGTACACAAATGAGTTTGTGATGAATAAAGTATGATGATTAAC 3657
Db 782 ---AspAspPheGlnGlnGlnGlnGln-----GlnGlnGlnAspGlnGlnGlnGln 798
QY 3658 GAAATATATGCTTTAGAGATTAACCAAGGACATGATGCTGTGCTGTTAAAGT 3717
Db 799 AlaGlnGlnValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 816
QY 3718 GATTCGAAACCGCAAGGATGATTAACCAAAAGCAAAAGCAAAAGCAAAAGCAAAAGTATAG 3777
Db 817 ProSerProProGlnThrThrGlnAspGlyValLysProAlaSerGlnGlnLysAspVal 836
```


OY	436	AGACGACGATGTCGATGATAAAACCTGGAAACCTGAATGATATAAAT---ACCCAAAT	432
Db	119	ATgLeuHhISValCysValATgAsnLeuLsInLsInLsAspTYrSerLysLLeAsnAsn	138
OY	493	ATTTCATGATTTTGTGGAAATCTACTACTGTACAGCAAAATTCGAAAGTCGAATCAATGTT	552
Db	139	LysHISAsnLeuLeuValGluValCysLeuLalLalLysTYrGluGluGluInsLLeIleThr	158
OY	553	AATTAATCATCCA---CATTAAGGAAC-T-----TCAGACGCTTGACTGCT	594
Db	159	GLyATgTYrTYrProGlnHISGlnGluLThrAsnProAspThrLysSerGlnLeuCysThrVal	178
OY	595	CTTGGACGAAAGTTTTCGAGATATAGTGATATTTGTGAAGAGAAATAGATGTTTAA---	651
Db	179	LeuAlaIArgSerPheAlaAspLeuGlyAspLLeIleArgGluLysAspLeuTYrArgGly	198
OY	652	-----CCAAATGTCATGCACAAAGTACAAACGGGCTCCGAGAGGTTTTC	696
Db	199	GLyAsnThrLysGluLysLysAlaArgLysLysLeuGluLsInLeuLysThrIlePhe	218
OY	697	AAGAAATACATGATGGAATG------GAGATGAAGTAAAAAT	735
Db	219	GLyHISLLeTYrAspGluLeuLysAsnGlyLysThrAsnGlyGluGluLeuGlnLys	238
OY	736	GATTACAAATCCGTATGATCTGGAATATTTATTAATTAAGAGAAGCATGGATGGATG	795
Db	239	ATgTYrATgAspLysAspAsnAspPheTYrGlnLeuArgLysAspTrpIArgPala	258
OY	796	AATTAACAAATTAAGTATGGGAGCATTAACATCAACGATGCA---TCATATAATCTGCATAT	852
Db	259	AsnATgGluThrValThrLysAlaLleThrCysAsnLalLeuTYrGlnLysGlnTYrSerGln	278
OY	853	TTTATGCATCAAGAAAGTAAATACCATATTTTTCAAATCTTAAGCGGCCATTAACAA	912
Db	279	ProThrCysGlyATgArgGlyGluLeuProTYrValThrLeuSerLysCysGlnCysIleAla	298
OY	913	GGAAGGTTCTTACCAAATTAGATATGTCCTCCCATATTTACGTTGGTTCCAGCAATGG	972
Db	299	GLyGluValProThrTYrPheAspTYrValProGlnTYrLeuATgTYrPheGluGluTrp	318
OY	973	GGAGAGAGTTTTCGCCAAAAAGAAATTAATTAATGAAAAAGTCGAAAGCATCCCTGCT	1033
Db	319	AlaGluAspPheCysArgGlyLysLysLysLLeProAsnValLysThrAsnCysArg	338
OY	1033	-----AATGCACAAAGACGCTTATATGTAGTCATATGACATGATTTGACGACA	1083
Db	339	GlnValGlnATgGlyLysGlnLys---TYCysAspArgAspLysTYrAsnCysAspGly	357
OY	1084	ACTATTTGGAAAAAGGATTTTTCGATTTGGATTAATAGTCATGACGTCCTGCATAA	1143
Db	358	ThrIleATgLysGlnTYrLLeTYrArgLeuAspThrAspCysThrLysCysSerLeuAla	377
OY	1144	TGCAAGTTTTCGAAAGTTTGCTTAAGGAAACAACAAGACGATTTAAAAACAAAAAGAA	1203
Db	378	CysLysThrPheAlaGluTrpLLeAspAsnGlnLysGluGlnPheAspLysGlnLysGln	397
OY	1204	AAATATGAAAAAGAAATCAATCATATTTATGCAACGATAAACAATTTGTCATATATTT	1263
Db	398	LysTYrGlnAsnGluLLeSerGlyGlyGlyATgArgArgGlnLysArgSerThrHISSer	417
OY	1264	AATATGAA-----TATTATTAACAATTTTATGAAGAAACTTAAGGAAAGCGCAATAT	1314
Db	418	ThrLysGluTYrGluGluTYrGluLysHISAsnGlnGluLeuArg-----	433
OY	1315	GCAACTATATGAC-----ACTTTTAAATTTTACTATTAACAGGAAAGTAT	1359
Db	434	-----AsnGlnLysAspValATgSerPheLeuGlnLeuLeuSerLysGluLysIle	451
OY	1360	TGTAAAGAGGATTTACCA---GGAGAAAAAGGATTTACTTCTTACATAAGTCGATGAC	1416
Db	452	CysLysGluATgGlnGlnValGlyGluGlnLysThrAlaAsnTYrGlyAsnProGluAsnGlu	471

QY	1417	AAAGGATATTTTATACGTTACGAAATATTCGCAAACTGTCGCCAGCGGGGGCAAACT	1476
Db	472	SeRanThrPheSerHisThrGluTyrCysAspArgCysProLeucGlyValAspCys	491
QY	1477	GATGTATATAAATACACACACAAATCAGATATGATGCTGACGCTGAATATATAGAAC	1536
Db	492	-----SerSerAspAsn-----	495
QY	1537	TATTAACCTCCATGGGGGTGTAAGCCCTACTAATATCACTGTCCTTATATAGTGTATGAA	1596
Db	495	-----	495
QY	1597	CAAGGTGATATTATACCAAAAAATTGAAAAATTTTGTACAGCTCACTAATTACAAAGAT	1656
Db	495	-----	495
QY	1657	AAAAATATCAAAAAATGGGAATGCTATTATTAAGAGTGAATAATTAATATAGATGAAC	1716
Db	496	-----	498
		CysProPhe	
QY	1717	GAACAAAAATCTGAAATCAATATGATATATCTTACATATATATCATTCATATATTTT	1776
Db	499	+++	499
QY	1777	GAATTATGGGTTACATATTATTTAAGGATACATATTAAGTGAATGCACAACTTAAAACT	1836
Db	500		507
		TriPalatrh	
QY	1837	TGTATATAATATATACAC-	1881
		ACCAGTTGATTTGATGATGATATACGA	
Db	508	CysThrGlyGlyThrAlaGlyThrGlyGlyGlyCysGlyAlaThrAlaGlyThr	527
QY	1882	AATGCTTATGTTTGGACAGTGGGTTAACAAAAAGAAAGATGATATATTAAG	1941
		ThrCysAlaCys	
Db	528	-----	531
QY	1942	AAACGTTCACAAAAAAGAAATATACAGCAATCGTATTATAGTATATTAATATATCTT	2001
Db	531	-----	531
QY	2002	TTTGAAGGTATTTTTTAAAGTTATGATTAACCTTGACAAAGATGAACAAAAATGAAA	2061
Db	531	-----	531
QY	2062	GAACCTTATGAAAAATATAAAAAGAAAAAATGACTTTCATTTGAAATATATAGG	2121
Db	531	-----	531
QY	2122	GACATTTAGAGATGCAATAGAAGACTCTTTAGATCACTTAAAGAAACGCCACATA	2181
Db	531	-----	531
QY	2182	TGTAAAGCAATTAATCAAAAGAGAGATGTGAACATCCCATTAATGCACAAACAAACCG	2241
Db	532	-----TriPalatrh	544
		GlyGlyGlyAlaThrGlyAlaAlaAlaGlyThrGlyThr	
QY	2242	TGTGTTAACTGCTGGAGCGACCAACCCATTAATAATATTAAGAAATATGCAACATAC	2301
Db	545	CysAlaAlaAlaGlyAlaThrThrTriPalatrh	555
QY	2302	TTTAAAGAGAGTGCATACGAGAAAGCAGCAAAATCGTGTTCATTAATTAATGAAGAAAG	2361
Db	555	-----	555
QY	2362	GCACACAGAGTATATATTAACGTGGGGGTAGAGAAAGACTTCAGAGCAATTAATGT	2421
Db	555	-----	555
QY	2422	AGATTAATGATTAACATCTATCTATCTTGTGTTTTCAAATGGACATGATGATGCG	2481
Db	556	-----ThrThrGlyAlaThrAlaGly	562
QY	2482	AAAGGACAGGTGATATACAAACAAAGATTTGCTAGAGAACTGATGGGAATGAT	2541

Db	563	AlaAlaThrAlaGlyGlyGlyAlaAlaGlyAlaAlaAlaGlyThr		577
OY	2542	CCGGACACATCGCTAAAGATCACGAAGATGTTATTATGCTCTCTAGAAAGACACATATA		2601
Db	577	-----		577
OY	2602	TGTACATCCATTGGACATTTNCAAAAGGATGATCACCCACTTAATGTAATATTGTT		2661
Db	577	-----		577
OY	2662	GATGATTAGTTAATAATTCCTTTTGGGGGATGTTCTTCTATCAGCAAAATATGAAGCA		2721
Db	577	-----		577
OY	2722	AACAAGATTAATACGATGTATAAAGAAAGAAATTAACCTTAAGGGCCCAAGAAATTA		2781
Db	577	-----		577
OY	2782	GACCCAAACACACGACACTATCTCCAGCTATACGTTACGTTTGCAGATATAGCT		2841
Db	578	-----		589
OY	2842	GATATAATTCGAGGAGAGATCTCTGGAAAGAAACGGTACATGTAAGCTGCAAGGA		2901
Db	589	-----		589
OY	2902	CATTGGAAACGTGTTTGGTAAATATACATTAAGTCACTCAAGGCAAGAAATGATATA		2961
Db	589	-----		589
OY	2962	TATATAATGATGCCCCCAATATTTTAAATTGAGGAAATTTGGTGGAAAGCTATATGA		3021
Db	590	-----		592
OY	3022	GCCAAAGTATGGAGGACATGTAATGATATATAATTTAAAGATTAATTCGGGAC		3081
Db	593	AlaAlaAlaGlyAlaThrThrGlyCys		601
OY	3082	CAATCAACACAAAGTAGTATATTCGGATATATGATCATACACCATGGATGATTAATATC		3141
Db	602	-----		607
OY	3142	CCACAAAAATTAAGATGATGACCGAATGGGCAATGCTACTGCAAGGTGCAGAAAA		3201
Db	607	-----		607
OY	3202	GAGTATGATAAGTTGAAGGAAAGTAAAGGTAAAGATTAAGATTAATGTCACAGC		3261
Db	607	-----		607
OY	3262	TGTACGAAAGAGAGTGTACAGCTTTGTACGAAGTGACAGAGCTTGTAAATGAATATAT		3321
Db	608	-----		619
OY	3322	GATATAATAGGATTGAGAAAGAACATGAATATATATATCAGATAAATACAAAGATTA		3381
Db	619	-----		619
OY	3382	CATGACACAGCAAAATGTCTGTACTAATATAGTGATTAAGAGTTCACAGTACAGCCAA		3441
Db	620	-----		634
OY	3442	AATCATATAGACAGAGATGTTATGAATTTTCTCGAATATTATACCAACAAATATGTGTC		3501
Db	634	-----		634
OY	3502	AAAAATATAAAAGTGTACTAGTATGAAGAGTGTCTGTCATTGGTACTACACACGAT		3561
Db	635	-----		636
OY	3562	GAATATGTGGAGCATCTCTCATGATACAGAAATTTGATGATGTCAGTCACACAAAT		3621

Db	637	-----	-----	-----	CysAlaIaIa	-----	639
OY	3622	GAGTTTGTGATGAAAAAAAGT	GATGGTAAAGATTAACGAAAAATATG	CTTTAGAGATPAA	3681		
Db	639	-----	-----	-----	-----	-----	639
OY	3682	CCACAGAGCCATGATG	TGTCGTGCTGGTGTGTTAAATG	TGATCGAAACGACAAAGG	741		
Db	640	-----	-----	-----	-----	-----	653
OY	3742	ATPAAAAAGAAAAAAGCGG	AGAAAGAAAGATACGAAATGAAACGTCGAATG	ATATA	3801		
Db	654	AlaIaIaThrGlyAlaThrIa	-----	-----	AlaIaIaCysAlaThr	-----	665
OY	3802	CTTAAAGAAAAAGATG	GAAGAAAGAAACAAGTAAAGATTTG	ATCCAAAAAGAAATAGTAA	3861		
Db	665	-----	-----	-----	-----	-----	665
OY	3882	GGATATCCCGATTGG	CAATCCGGAATTAATTAATTA	TAGTGAAGACCCTCGTGTATG	3921		
Db	666	-----	-----	-----	-----	-----	668
OY	3922	CCCCCTGAAGACAAAG	TGATCGTACATTTCTTGCCAAATG	ATAAATGAAATAAAAA	3961		
Db	669	GlyIaIaIaGlyIaIaThrCys	-----	-----	-----	-----	676
OY	3982	TTACATACAGATTA	TTTTAAAAAGCTTTCATCAATTCG	CAGCAGCAAAACATTC	4041		
Db	677	-----	-----	-----	-----	-----	686
OY	4042	TTTCATGCTATTATT	TTAAAGTAAAGATGGTGAAGCAATGA	ACTCGAATAAAGAA	4101		
Db	686	-----	-----	-----	-----	-----	686
OY	4102	AAAGAAGCAAAATTC	CTCCCGATTTTGAGATCATGTTCT	ACATTTGGAGATTAT	4161		
Db	686	-----	-----	-----	-----	-----	686
OY	4162	AGAGATTTTTTATTTG	GAACAGATATATCAAAAGTCATGGTGA	AGCAATGAACTPAAA	4221		
Db	687	-----	-----	-----	-----	-----	698
OY	4222	GAGCAAAATAGATTCT	CTTTTCAAAAATGGTGACCAAAAATCT	CTAATGAAAAACAGC	4281		
Db	698	-----	-----	-----	-----	-----	698
OY	4282	CAAGAATGCTGAGAC	AGAACTACTCATGATATGGAAGCTATGCT	TGCGACTAGTA	4341		
Db	699	-----	-----	-----	-----	-----	705
OY	4342	AAAATTTGGGCAAAAA	AGATGATTTTACCGAAACTACGCTTAC	ACAACGTCMAATTT	4401		
Db	706	AlaIaThrGlyIa	-----	-----	-----	-----	720
OY	4402	AGTACAAAGACACC	ACTTTGGAGAAATTTGCCAAACGACCC	CACTTTTACATGGCTA	4461		
Db	720	aAlaIaThr	-----	-----	-----	-----	722
OY	4462	ACCGAATGTCAGCAG	CACTTGTCTATACAGCAAAAAATTTT	CAAGATGTCAGAA	4521		
Db	723	-----	-----	-----	-----	-----	727
OY	4522	AAATGTAAGTCAATG	ACCAATGGAAGTGTATACAGAAATGTAAT	TAAAGAAATGCGAGAG	4581		
Db	728	ThrCysAlaThrGlyIa	ThrThrThrThrThrGlyIaCys	-----	-----	-----	741
OY	4582	TACGTTAAATATATG	AAAAAAGATGAGATTC	CAAGATTAATTTACAAGAT	4641		
Db	741	-----	-----	-----	-----	-----	741
OY	4642	GAACGCAAAAAAAGAT	TGCATACACACACATTTGCTGTAATG	CTTACAGACTATACT	4701		
Db	742	-----	-----	-----	-----	-----	753

Db 1022 SerAsnValThrPheAsnGlyHisLeuThrGlyThrLysLeuAsnGluPheAlaSerArg 1041
QY 6904 CCCCAATTTTGGAGTGGCAGAGAAATGGGCAAGAAATTTTGTATATAGAGAAAGAA 6963
Db 1042 ProSerPheLeuArgThrPheThrGluThrGlyAspGlnPheCysArgGluThr 1061
QY 6964 CAGTTGTTAAATTTGAGCGGCGCTGAAGAAATATGATGTAATGCT-----AGTAAT 7017
Db 1062 GlnLeuGlnIleLeuLysGluArgCysMetValTyrGlnThrAsnGlyAspLysGly 1081
QY 7018 GAGGTTAGACACAGACATGTGCAGCGCTGTGTACATATCAAAATTTTATTAAAGAG 7077
Db 1082 AspAspLysLysGluLysCysThrGluAlaCysThrTyrTyrLysGluThrLeuThrAsn 1101
QY 7078 TGGAAACCTGATATGTGAAACAAAGAAAGTCAAAAGTAAAGTATGCAAAAGAG 7137
Db 1102 TyrGlnAspAsnThrLysLysGlnAsnGlnArgTyrThrGluValLysGlyThrSerPro 1121
QY 7138 TATAGGATTATCTCTTCTACTGAAAGACATAGACAGCAACATGTGCTCATGATAT 7197
Db 1122 TyrLysGlu-----AspSerAspValLysGluSerLysTyrAlaHisGlyTyr 1137
QY 7198 TTTAACTGAAATTTAAAGAA---TATCTGCGCAATAGAGAT-----TGTCT 7242
Db 1138 LeuArgLysIleLeuLysAsnIleIleCysThrSerGlyThrAspIleAlaTyrCysAsn 1157
QY 7243 TGTATGCAAAACCTTCTTCTACACACTACCAAAACAAACACACATCCATCCGAT 7302
Db 1158 CysMetGlu-----GlyThrSerThrAspSerSerAsn 1169
QY 7303 GCTAATGATATGCGCAGATGCGATATATGTTCTGCAAGAAATTTTAAACAG--TGTGAG 7359
Db 1170 AspAspAsnIleLeuProLysLeuLysTyrProProIleGluGluGlyCysThr 1189
QY 7360 TGTCTGCACTTTCAAAAGGATCTATGATTCATACAAAAATTTACTGACCTTAA 7419
Db 1190 CysLysAspProSer---ProGlyGluValIleProGlyLysValIleProGluProLys 1208
QY 7420 ATACCTATGATGTTGTGAGAAAGCAGCATTTATTTATCTAAAGAACAGAAATTAAT 7479
Db 1209 Val----- 1209
QY 7480 ATGATATTTACCTTGAGAAATAATTTATCTATGAGTCTACAAAGAAAGAAAGT 7539
Db 1209 ----- 1209
QY 7540 AAAAATAGTTGACATAATATATCTTGCGATCTTACAGAACTTATGACCTGATATA 7599
Db 1209 ----- 1209
QY 7600 TATATAGAGAGAAACCTTGTGAAATATAGAGAAATATCGTTTAAAGTAGATTAAT 7659
Db 1209 ----- 1209
QY 7660 GAATGCAATGTTACAAAATTCAAAGTCTATACGAGAGAAAAAGAGATGTGTACT 7719
Db 1209 ----- 1209
QY 7720 CCAAGAGAGACATATGTCTTAAAGAAATTTAGATCAATTTAAATTTGAACACTTAAG 7779
Db 1209 ----- 1209
QY 7780 GATAGTATATCTCTCAAAATATGTTCTGCAACTGCAGAAATGAAAGAAATAGACATA 7839
Db 1209 ----- 1209
QY 7840 ATAAAAAATTCAACTCAGAGAAAGGCTGCGCAATGCAATATGATGATCTATACAAA 7899
Db 1209 ----- 1209
QY 7900 TATAGTTTCGCTGATCTGGGTGACATAGTTAGAGAAACAGATATGTTACGAATTTGCTGT 7959

Db 1209 ----- 1209
QY 7960 TACTTACCTCCGTAGAAATTAATATATATAGGTTTTGATATACATATATGAAAAATGG 8019
Db 1209 ----- 1209
QY 8020 AGAAATTAATAAAGTAGAAATTAATAACACAGATGTACAAACGTTCTGCTGTGG 8079
Db 1209 ----- 1209
QY 8080 TGGGATGCTATATGAAAAAGATATTTGAAAGCAATGACGTGCAAACACCAGAGATGCA 8139
Db 1209 ----- 1209
QY 8140 AAACTTTATAGAAAGAGAAATGATGATTTGAACGCAATACATTAATACAGATTAAG 8199
Db 1209 ----- 1209
QY 8200 TGTGACATTAAGAGCATGCACCTGTGATGATTAATATACCTCAACGTTTCGATGAGAG 8259
Db 1209 ----- 1209
QY 8260 ACTGAATGCTGATATATTTATTTAAAGCAGCTAGTGAGAAATTTGAAAAATTTAAAAA 8319
Db 1209 ----- 1209
QY 8320 TCATGTGATCATCTGTAAACATCTGACAGATGCAGAAATGATTAATGAAATAAGTGT 8379
Db 1209 ----- 1209
QY 8380 GAACAGTGTAAACGAGATGTCAGAAATATTAATAATTTGTTCTTAATGGAATCTCTA 8439
Db 1209 ----- 1209
QY 8440 TTGATATATACATCAATTAATATCAAAAGATTTATAGAACACCAATATATACAAAAATC 8499
Db 1209 ----- 1209
QY 8500 TCTACTTATGATCATGTCTCAAAATTTTGTACAAAGTGTGAAACCTTTAAAGTGAATGT 8559
Db 1209 ----- 1209
QY 8560 TCTGTTGAGAGCTTTCTGCAATATCTTCATGAAACAAAGTAGTGTGAAATTAATTAATTT 8619
Db 1209 ----- 1209
QY 8620 AATGAAATGATGTCTTCCATATACGAAACATATGCTTTGAGAAACACCAAAAAGT 8679
Db 1209 ----- 1209
QY 8680 TATTAAGAAAGCTTGCAATGTACACTACTCTTAAGAAATCATGTGATTAATGTCTTACC 8739
Db 1210 -----LeuProLys----- 1212
QY 8740 GATCAAAACAAAGATGAGTAAGAAATTAACAACTTTACCTTCGTCGAGAAATGAT 8799
Db 1212 ----- 1212
QY 8800 TATGATATATATCTGTATATTTGAAACGCATACCTTGTCTTAATAGTTCAAGATGAAC 8859
Db 1212 ----- 1212
QY 8860 AAAGTGATATGATCTCCCAAGAAAGACATTTATGTCAAGACCTATACACTGCATAT 8919
Db 1213 -----ProLys----- 1215
QY 8920 AATTATAGAAAAAGGTATTAAGAAATTTTAAAAAAAACCTTCTTACTCTGCTTCAAGT 8979
Db 1215 ----- 1215
QY 8980 CAAGACAATTTGTTAGGTCAAAAAATATTAATCGAAGAGAGTGTGCTTTGAGGCAATG 9039
Db 1215 ----- 1215

QY 9040 AATATAGTTATGAGATTAATCCGATTAATTAAGAACTGATATGAGACTTCA 9099
 Db 1215 -----
 QY 9100 TATCTGAAAAATTAATAAATTTATTAACATCAATGAAGCAACCGAAATCGTAA 9159
 Db 1215 -----
 QY 9160 ACATGTGGGAAATATAGAGTCAGATATGGCAGCTATGATATGATATAAATT 9219
 Db 1215 -----
 QY 9220 GCTACTCAAAAGTAACATTAAGTAGAGAGATGTCATTAACCAAGATGAAGAACT 9279
 Db 1216 -----
 QY 9280 AATCAGTTCTTCTGTGTTAATTAATGAGGCAAGCAAGCATGTAAGAAAGAAACAT 9339
 Db 1218 -----
 QY 9340 GTAAAGTATTCATTAATAAACAATAATGCTTCGTTCAACAGAAATATTTGAAGCTCA 9399
 Db 1218 -----
 QY 9400 GAATTAATTAAGCAACCTGATGTCAGAAATGATATTAAGAAATATTAATTAATTA 9459
 Db 1219 -----
 QY 9460 TTGATAAAAATCAATGAGAAATCTAATATATAAATATAGCAATTAAGATCAATC 9519
 Db 1221 -----
 QY 9520 TCAGTAATATAGCAATTAACCATCTGAAGAAATGTCAGTCAATATTAATCAAAA 9579
 Db 1222 -----
 QY 9580 GATTCATATGCGCTTGGAGTTAATGATTAATTAATGAATAGTACAGAACCAAAAAT 9639
 Db 1225 Asp -----
 QY 9640 AATGAAAAATATGAAATCAAGAGTACTAAAAAATATATTCGTTATATTTTGT 9699
 Db 1225 -----
 QY 9700 GAAGATGAAGAACACAAAAATCATGACTAGATGAATATTAAGAACAGAGCAACA 9759
 Db 1225 -----
 QY 9760 GTTCGTCCTAAGCACTATTTCTTTACACCCCATGATGTTCTTCTATCAAGACCT 9819
 Db 1226 -----
 QY 9820 TTAATCTCAACACATGAGTACACATATGATCTCTAAATGATATATGAAAAAGT 9879
 Db 1232 -----
 QY 9880 ATCTCTGTTGTTATGATCGCGCTAGGTTGATAGCGCTCATTTGATGAAGAAAA 9939
 Db 1240 IleMetIrpSerIleGlyLeuIleAlaIlePheIleIlePheIleIleIleIleIle 1259
 QY 9940 TTCAAATGCTGTCGAGCTGTCGTAATGATGAAATATCCGCAAGAGAGATGATGATG 9999
 Db 1260 ThrLysSerThrIleAspLeuIleValIleAsnIleProIlePheIleIleIleIle 1279
 QY 10000 CCTAGTGAATCCAAATAGTACATACATTAATGAAATGTCCTATTAAGCAAA 10059
 Db 1280 ProIleLysLeuSerProAsnArgIlePheProIleIleIleIleIleIleIleIle 1299
 QY 10060 ACATATATATATGAGAGAGATACT -----AGTGAAGATGAAGATATATATG 10110
 Db 1300 ArgTyrIleTyrLeuGluGlyAspSerGlyThrAspSerGlyTyrThrAspHisTyr 1318

QY 10111 TGGACTTATCTCTCTGATTAATCTACT-----TCATCCGAAAGTAGATGAAGATTTGAT 10167
 Db 1319 -----
 QY 10168 ATTAATGATATATATGATACAGTACTCTCTAAATATTAAGATGATGATGATGATGAT 10227
 Db 1334 IleAsnAspIleTyrAlaProIleGalaProIleIleIleIleIleIleIleIleIle 1353
 QY 10228 GAACATCAAAAAAGGATTTACCAAGT-----GATGATACCAAGT----- 10269
 Db 1354 GluProSerGlyAsnAsnThrThrAlaSerGlyAsnAsnThrProSerAspThrGlnAsn 1373
 QY 10270 -----AATGATACACACAGTACATGATTAATGATGATGATGATGATGATGAT 10320
 Db 1374 AspIleGlnAsnAspGlyIleProSerSerIleIleIleIleIleIleIleIleIle 1393
 QY 10321 AAACATGATTTGTTATTCATATATTAATTAACCAATACAGAACCAAT-----AATAT 10371
 Db 1394 LysAspGluPheIleSerGlnTyrLeuGlnSerGlnProAsnThrLeuIleIleIleIle 1413
 QY 10372 TCAAAAGTGCAGATATTCATGATGATGATGATGATGATGATGATGATGATGAT 10431
 Db 1414 TyrSerSerGlyAspIleProLeuAsnThrGlnProAsnThrLeuIleIleIleIle 1433
 QY 10432 GAAGAAAAACCTTTATATATATCTATTCATGATGATGATGATGATGATGATGATGAT 10491
 Db 1434 AspGluLysProPheIleIleIleIleIleIleIleIleIleIleIleIleIleIle 1453
 QY 10492 ACTTATATATATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 10551
 Db 1454 SerTyrAsnValAsnMet-----ValAsnThrAsnAsnAspIleProIleSerGlyLysAsn 1472
 QY 10552 GATTCCTTATAGAGTATGATGATTAATTAATGATGATGATGATGATGATGATGATGAT 10611
 Db 1473 GlyTyrTyrSerGlyIleAspLeuIleAsnAspSerLeu-----AsnSerAsnValAsp 1492
 QY 10612 TATATGATGATGATGAT 10628
 Db 1492 LeTyrAspGluValLeu 1497

RESULT 14
 AAR70235
 ID AAR70235 standard; Protein: 700 AA.
 AC AAR70235;
 DE 22-SEP-1995 (first entry)
 DE P. falciparum EBL-e2.
 KW Erythrocyte binding ligand; EBL-e2; binding domain; malaria; therapy;
 KW vaccine.
 OS Plasmodium falciparum.
 PN W09507353-A.
 PD 16-MAR-1995.
 PF 07-SEP-1994; 94WO-US10230.
 PR 10-SEP-1993; 93US-0119677.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI; 1995-123427/16.
 DR N-PSDB; AA083528.
 PT New erythrocyte binding domain polypeptide(s) - isolated from
 Plasmodium binding proteins, used in diagnosis, treatment and

PT prevention of malaria
 XX Disclousure; Page 56-57; 81pp; English.
 XX
 CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AA083526), E3la (AA083527), EBL-e2 (AA083528) and Proj3
 CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli.
 CC Yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 CC
 XX Sequence 700 AA:
 SQ
 Alignment Scores:
 Pred. No.: 3.23e-64 Length: 700
 Score: 1032.50 Matches: 247
 Percent Similarity: 50.288 Conservative: 110
 Local Similarity: 34.798 Mismatches: 250
 Indels: 103
 Gaps: 22
 Gap: 16
 US-10-087-013-1 (1-10628) x AAR70235 (1-700)
 QY 406 AATGATGCTAATGCTGTCGCGCCACCTAGACAGACATATGTGTAAATAAATTGGAA 465
 DB 5 Asnlyvalgllyalacysalaprotylrgdrglnehlislcysaspylraenleuglu 24
 QY 466 GCTCTAATGATTAATATCCCAAAATATTCATGATTTATGGGAATGTACTAGTTACA 525
 DB 25 Serilleasp-----ThrlrSerThrlrhlislyseuleuenglualcysmala 42
 QY 526 GCAAAATACAGAGGTGATCAATGTTAATAATCAT-----CCACATAAAGCACT--- 576
 DB 43 Alalyetyrcluelylanserle---Asnrlhlisetyrthrglnhlsglnatylthrasn 61
 QY 577 -----TCAGACCTTGTACTGCTCTGTCGACGAAAGTTTGCAGATATAGTGAT 624
 DB 62 Gluaspseralaserglneucystlvalleualalargserpheaalaasplileglyasp 81
 QY 625 ATTGTAGAGGAATAGATATGTTT-----AAACCAATGTCCATGAC 666
 DB 82 lilevalrgegllyllyaspleuylrleuglyltyraspsnlysglulysglulnarglys 101
 QY 667 AAGTAGAAGAGCGGTCTCCGAGAGGTTTTCAGAAATACATGATGGAATGGAAGATGAA 726
 DB 102 Lysleuglulnlylseuylasprllepheylsylehlis-----lysasprval 118
 QY 727 GTAAAAAATGAT-----TACATCTGATGCA---TCTGGAATTTATAT 768
 DB 119 Metlyerthrasnlyalaglnlualrgtyrllleaspralalalysgllyasprphe 138
 QY 769 AATTAAGAGACGATGCGATGATGGAATGAATTAAGTATGGAAGATGATACATGTC 828
 DB 139 GlhleaufgelnllyasprlthrlrSerasnaarglulthralilrlysalaleuilecys 158
 QY 829 GATGATCATATAATCTGATATTTATGATCAGAAAGTAAATACACCATATTATTTC 888
 DB 159 Hlsalalrprolyslualaaenlyrphelileylthralacysasnvalgllylsglythr 178
 QY 889 AATCTAATATGCGCCATTAACAAGAAAGCTTCTACCAATTTAGATTATGTCCCTCAA 948
 DB 179 Asngllylncyslhlscyslilegllyllyasprvalrprothrlrlyrphasprlyvalrprogl 198
 QY 949 TATTTACGTTGCTTGGAGGAGAGAGATTTTCCGAAAAAAGAAATATTAAATG 1008
 DB 199 lylleuaftrpnegluglultrpalaagluasprheycysllylsglylsglylsglyl 218
 QY 1009 AAAAGTCAAGACTCTGCTGTAATGACAAAGACGTTATATTATCTATCAATGAA 1068
 DB 219 Gluasnleuglnlylsglncyslarsgsplyrcluglnasneulyrcyssergllyasncl 238

QY 1069 CATGATTTGACAGCAACTATTTGGAAAAAGATATTTTGGATTTGGATTAATAGTACT 1128
 DB 239 TyrAspcyslhrlystrlletyrllylsglyllylevalillegllylunhlscystlhr 258
 QY 1129 GACTGTTGACTTAAATGCAAAAGTTTGGAAAGTTGGTGGAGATACACAGACATTT 1188
 DB 259 Asncyservalrprcysatgmetylrlylthrlpilleaspsnlnlylsglylsglyl 278
 QY 1189 AAAAACAACAAAGAAATATGTAAGAAATAT----- 1221
 DB 279 Leulysglulnlylsglyltyrlylthrglnlilesergllylsglylsergllylsglyl 298
 QY 1222 -----CAATCATATTTATCGAACGATTAACAATTTGTCAAT 1257
 DB 299 LysarglthlylsglalaalaargserSerSerSerSerAsprasn----- 314
 QY 1258 AATATTAATGTAATATTTTAAACAATTTATGAAAACTTAAGAAACGCAATATGCA 1317
 DB 315 -----gllylrglserllylspheyllylsglylsglylsglylsglylsglyl 330
 QY 1318 ACTAATGACACTTTTAAATTTACTTAATGAGAAAGATGTTGTAAGAA----- 1368
 DB 331 Asprvalasprlyspheleuylileleuasnllylsglyllyllecysglulnlylsgln 350
 QY 1369 ---GGATTACAGAGAAAAAGATATTACTTTTACTATACAGTCTGATGACAAAGGATA 1425
 DB 351 Valgllyasnclulysalaaasprvalasprhetlrasnclulysltyrvallys---Thr 369
 QY 1426 TTTTATGCTTCAATATTTGCCAAGTGTGCCGACTGGGGGTCAATGTGATGTATA 1485
 DB 370 PheSerarglthrglullecysgluprocyslprotyrcysgllyleuulysgllylpro 389
 QY 1486 AATTAACACACACAAATCATGATGATTCGCAACGTGTAATGATGATGATGATGAT 1545
 DB 390 Protyrlylvallylsglylsglylsglylsglylseralalyslthrltyrgrpro 409
 QY 1546 CCATGGGGTGTGAACCTTAATATCATCTGCTTTATAGTGTATGATGACAAAGGTGAT 1605
 DB 410 Lys-----Asnlllethraspilleprovalleuyltyrproasprlysserglnln 426
 QY 1606 ATTACACAAATTTAGAAATTTTGTACACGCTCACTAATTAACAAAGTATTAAT 1665
 DB 427 lileuulysllyltyrlylssasprheycysgllylsglylalaPro--gllylsglyl 445
 QY 1666 CAATAATGGAATGCTATTTAAGAT-----GAAATATTAATAGATG 1710
 DB 446 Lyslyltyrlnlncysltyrlylgrasglunhlslargproseterlylssnlnasln 465
 QY 1711 AATCTGGAACAAATCTGAAATCAATGATGATATCTAGATTAATATTCATTT 1770
 DB 466 Valgluglylthrlrprsprlylspheylnglnlylsglnlthrgllylssertylrasnval 485
 QY 1771 TTTTGAATTAATGATGATTTATTAAGGATGATGATGATGATGATGATGATGAT 1830
 DB 486 Phehetrprsprlylhlisaspmetleuinhlsasprvalclultrpulylsthrgl 505
 QY 1831 AAACTGTATAATTAATACACACG-----CATGTATGATGATGATGATGAT 1878
 DB 506 Setlylscyllelssnlnasnlthrasnnglylaspthrcyslrgsnlnasnlscyls 525
 QY 1879 AGAATTTGCTTATGCTTGCACAGATGGTTAAACAAAGACAGATGGAATGATATA 1938
 DB 526 ThrAspcylsglylspheglulstprvalglulyslelsglnlultrpmetalaile 545
 QY 1939 AAGAACTGTTCACAAAAAAGATATTCACACATCGATATATATATATATAT 1998
 DB 546 Lysasprhlsphecllylsglnlthrasprllevalglngln-----lyslyl 560
 QY 1999 CTTTTCGAAGTATTTTAAAGTATGATGATGATGATGATGATGATGATGATGATGAT 2058
 DB 561 LeulilevalpbeserProtyrlylvalleuasprleval-----Leulysglly 576
 QY 2059 AAGCACTTATGCAAAATATTAACAAAGAAAAAATGAGTTTCCAAATTTGGAATAT 2118

```

Db 577 GlyAsnLeuLeuGlnAsnLeuLeuSAspValHisGlyAspThrAspAspIleuHisHisle 596
Oy 2119 AGGACATATTAGAG-----::: :::::::::::
Db 597 LysLysLeuLeuAspGluGlnAspAlaValAlaValleuGlyGlyLysAspAsnThr 616
Oy 2137 GCATATGAACTTGTATGATCATCTTAAAGAACTGCCACGATATGTAAAGACATATAT 2196
Db 617 ThrIleAspLysLeuLeuGlnHisGlnLysGluGlnAlaGluGlnCysLysGlnLys--- 635
Oy 2197 ACAACGAAAGCATGTGA-----ACATCCCATATATCAACACACA 2235
Db 636 ---GlnGluGlnLysCysGlnLysLysAlaGlnGlnLysSerArgGlyArgSerAlaGluThr 654
Oy 2236 AACCCGTCGTATTAACCTCGTCGAGCAGCACCACCACTAAATATATAAGAAATAGCA 2295
Db 655 ArgGluAspGluArgThrGln-----GlnProAlaAspSerAlaGlyGluValGlu 671
2296 CAAATCTTAAAGAGATGCATACGAGGAA 2325
:::
672 GlnGluGlnAspAspAspAspTyrAspGlu 681

```

RESULT 15

AAW22481 ID AAW22481 standard; Protein: 700 AA.

AAW22481; AC

07-OCT-1997 (first entry) DT

Plasmodium ebl-2. DE

DBL gene family; SAMP; sialic acid binding protein; vaccine; therapy; KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte; DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response; Plasmodium. KW

Plasmodium falciparum. OS

MO9640766-A2. PN

19-DEC-1996. PD

07-JUN-1996; 96WO-US09508. PF

07-JUN-1995; 95US-0487826. PR

(USSH) US DEPT HEALTH & HUMAN SERVICES. XX

Chitnals C, Miller LH, Peterson DS, Sim KL, Su X; Wellens TE; PI

WPI: 1997-052231/05. DR

N-PSDB; AAT72896. DR

New malaria vaccines - contains cysteine-rich DBL family protein PT

binding domains homologous domains of the Duffy and sialic acid PT

binding proteins XX

Disclosure; Page 46-48; 96pp; English. PS

This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to CC

the Duffy binding like (DBL) family of genes which have homology to CC

Duffy antigen binding protein (DABP) and sialic acid binding protein CC

(SABP) conserved regions (see AAT72889 and AAT72888 respectively). The CC

var family of genes modulate cytoadherence and antigenic variation of CC

Plasmodium infected erythrocytes. SAMP and the Duffy antigen binding CC

protein (DABP) are soluble proteins that appear in the culture CC

supernatant after infected erythrocytes release merozoites. DABP and CC

SAMP mediate the binding of merozoites and schizonts to the erythrocyte CC

surface. These proteins are necessary for erythrocyte invasion by the CC

parasite. This sequence can be used in the compositions of the CC

invention. The compositions are for the treatment and prevention of CC malaria, and comprise either a nucleotide sequence or encoded polypeptide CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a CC family of genes having homology with conserved regions of DABP and SAMP. CC The compositions are used for the treatment and prevention of malaria. CC They are also used in the preparation of vaccines for inducing a CC protective immune response in a mammal to Plasmodium merozoites CC (especially Plasmodium falciparum or Plasmodium vivax).

Sequence 700 AA;

Alignment Scores:

Pred. No.:	3,236-64	Length:	700
Score:	1032.50	Matches:	247
Percent Similarity:	50.28%	Conservative:	110
Best Local Similarity:	34.79%	Mismatches:	250
Query Match:	5.31%	Indels:	103
DB:	18	Gaps:	22

US-10-087-013-1 (1-10628) x AAW22481 (1-700)

```

Oy 406 AATGATGCTATAGCCTGTGGCCACCTAGAGACGACATATGTGATAAACTGGAA 465
Db 5 AsnLysValGlyAlaCysAlaProTyrArgArgLeuHisLeuCysAspTyrAsnLeuGlu 24
Oy 466 GCTCTAAATGATATTAATACCCAAATATTCATGATTTATTTGGAAATGATCTAGTTACA 525
Db 25 SerIleAsp-----ThrThrSerThrThrHisLysLeuLeuGluValCysMetAla 42
Oy 526 GCMAATACGAAAGTGATCAATGTTATATATCAT-----CCACATAAAGAACT--- 576
Db 43 AlaLysTyrGlnGlnLysAsnSerIle---AsnThrHisTyrThrGlnHisGlnArgThrAsn 61
Oy 577 -----TCAGACGCTGTCTGCTGCTGTGACGACAGGTTTGGATATAGTGGTAT 624
Db 62 GluAspSerAlaSerGlnLeuCysThrValLeuAlaArgSerPheAlaAspIleGlyAsp 81
Oy 625 ATGTAGAGGAAATAGATATGTT-----AAACCAATGTCCATGAC 666
Db 82 IleValArgGlyLysAspLeuTyrLeuGlyTyrAspAsnLysGlnGluArgLys 101
Oy 667 AAAGTGAAGACGGGTCTCCGAGAGGTTTCAAGAAATACATGATGATGAAGATGAA 726
Db 102 LysLeuGlnGlnLysLeuLysAspIlePheLysLysIleHis-----LysAspVal 118
Oy 727 GTAAAAAATGAT-----TCAATCCTGATGGA---TCTGAAATTTATAT 768
Db 119 MetLysThrAsnGlyAlaGlnGluArgTyrIleAspAspAlaLysGlyLysPheHe 138
Oy 769 AAATTAAGAGAGCATGTGGATGTGAATGAATAAATAGTATGGAGCATTAACATGT 828
Db 139 GlnLeuArgGluAspTyrThrPheSerAsnArgLysThrValThrLysAlaLeuIleCys 158
Oy 829 GATGATCATATTAATCGATATTTTATGCAATACAGAAATATATACCCATATTTTCA 888
Db 159 HisAlaProLysGlnAlaAsnTyrPheIleLysThrAlaCysAsnValGlyLysGlyThr 178
Oy 889 AATCTTAATATCGCCGACATAAACAAGAAAGCTCTCAATTAATAGATTATAGTCCCTCA 948
Db 179 AsnGlyGlnCysHisCysIleGlyGlyAspValProThrTyrPheAspTyrValProGln 198
Oy 949 TATTACGTTGTTGCGAGAAATGGGGAGAGATTTTCCGAAAAAGAAATATTAATTTG 1008
Db 199 TyrLeuArgTyrPheGlnGluThrAlaGlnLysPheCysArgLysLysLysLeu 218
Oy 1009 AAAAAGTCAAGACACTCTCTGCTGATATACAAAGACCTATATTTGATGATCAATATGGA 1068
Db 219 GluAsnLeuGlnLysGlnCysArgAspTyrGlnGlnAsnLeuTyrCysSerGlyAsnGly 238
Oy 1069 CATGATGTAGACAACTATTGTAAGAAAGCTATTTGATTTGGAATTAATAGTACT 1128
Db 239 TyrAspCysThrLysThrIleTyrLysLysGlyLysLeuValIleGlyGlnHisCysThr 258

```

```

QY 1129 GAGGTTCGACTAAATGCAAGTTTTTGAAGTTGCTTAGCAATGACAGAACATTT 1188
Db 259 AsnCySerValrPrpCysArgMetYrGluThrPileAspAsnGlnLysGluPhe 278
QY 1189 AAAAAACAAAAAGAAATATGAAAAAGAAATA----- 1221
Db 279 LeuLysGlnLysArgLysYrGluThrGluLileSerGlyLysGlyLysSerPro 298
QY 1222 -----CAATCAATTTATTCGACAGATTAACAAATTTGTCAT 1257
Db 299 LysArgThrLysArgAlaAlaArgSerSerSerSerSerSerSerSerSerSer 314
QY 1258 AATATTTATAGTGAATTTTAAACAATTTATGAAAAAATTAAAGAAAGCAATATGCA 1317
Db 315 -----GlyYrGluSerLysPheYrLysLysLeuLysGluValGlyYrGln 330
QY 1318 ACTAATGACACTTTTAAATTTACTAAATGAGAAAGATTTGTTAAAGCA----- 1368
Db 331 AspValAspLysPheLeuLysLileLeuAsnLysGluGlyLleCysGlnLysGlnProGln 350
QY 1369 ---GATTTACAGAGAGAAAGATATTACTTTACTAACAGTGCCTGATGACAAAGGATA 1425
Db 351 ValGlyAsnGlnLysAlaAspAsnValAspPheThrAsnGlnLysYrValLys---Thr 369
QY 1426 TTTTATGCTTCAGAAATTTGCCAAGTGTGTCCGACTGCGGGTCAAAATGTATGCTATA 1485
Db 370 PheSerArgThrGluLleCysGlnProCysProThrProLysGlyLeuGlnLysGlyPro 389
QY 1486 AATPACACACAAATCAGATTAATGATCGAAGTGTAAATGATGAGACTTAAACCT 1545
Db 390 ProThrLysValLysGlnAspLysThrCysGlySerAlaLysThrLysThrYrAspPro 409
QY 1546 CCATGGGGTGAACCTCTAATATCATCTGCTTTAGTGTAGTAAAGCAAGGTGAT 1605
Db 410 Lys-----AsnLleThrAspLleProValLeuYrProAspLysSerGlnGlnAsn 426
QY 1606 ATTACACAAAATTAAGAAATTTTGTACAGCTCAACTAATTAACAAGATAAATAAT 1665
Db 427 LileuLysLysYrLysAsnPheCysGlnLysGlyAlaPro---GlyGlyGlnLile 445
QY 1666 CAAAAATGGAATGCTTTTATTAAGAT-----GAAATATTAATAGATGT 1710
Db 446 LysLysThrProGlnCysYrGlyYrAspGlnHisArgProSerSerLysAsnAsnAsnCys 465
QY 1711 AAACGTGGAACAAATACGTAATCAATTAATCAATTAATCAATTTTCATTAAT 1770
Db 466 ValGlnGlyThrTrpAspLysPheThrGlnGlyLysGlnThrValLysSerYrAsnVal 485
QY 1771 TTTTTCATTTATGCTTACATATTTATTAAGGATCTATTAAAGTGAATGACAAACTT 1830
Db 486 PheSerThrAspTrpValHisAspMetLeuHisAspSerValGluTrpLysThrGlnLeu 505
QY 1831 AAACCTGTATTAATATATACACACG-----CATGTATGATGAATGTAAAC 1878
Db 506 SerLysCysLileAsnAsnThrAsnGlyAsnThrCysArgAsnAsnAsnLysCysLys 525
QY 1879 AGAAATGCTTATGCTTTTGACAGATGGGTTAAACAAAAAGAAAGAAATAGTATA 1938
Db 526 ThrAspCysGlyCysPheGlnLysTrpValGlnLysGlnGlnGlnLutrmcAlaLile 545
QY 1939 AAGAAACTGTTCACAAAAAAGAAATATACAGCAATCTATTATAGTAATATTAAT 1998
Db 546 LysAspHisPheGlyLysGlnThrAspLileValGlnGln-----LysGly 560
QY 1999 CTTTTCAGAGTTATTTTAAAGTATGATTAATCAACTGACAAAGATGAAGCAAAATG 2058
Db 561 LeuLileValPheSerProYrGlyValLeuAspLileVal-----LeuLysGly 576
QY 2059 AAAGAACTTATGAAAAATATAAAAAGAAAAAATGACTTTTCCAAATTTGAAATTAAT 2118
Db 577 GlyAsnLeuGlnAsnLileLysAspValHisGlyAspThrAspAspLileLysHisLile 596
QY 2119 AGGACTATTTAGAG-----AAT 2136

```

```

Db 597 LysLysLeuLeuAspGlnGlnLysAlaValAlaValValLeuGlyGlyLysAspAsnThr 616
QY 2137 GCAATGAACTCTCTGTAGATCACTTAAAGAAACTGCCACGATATGTAAGCAATTAAT 2196
Db 617 ThrLileAspLysLeuLeuGlnHisGlnLysGlnAlaGlnGlnLysGlnLys---- 635
QY 2197 ACAACGAAAGCATGTGAA-----ACATCCCAATAATGCAACACAA 2235
Db 636 ---GlnGlnGlnCysGlnLysLysAlaGlnGlnLysSerArgLysArgSerAlaGlnThr 654
QY 2236 AACCCGTGTCTTAACCTGCTGAGGACGCCACCTAAATATTAATAAGAAATAGCA 2295
Db 655 ArgGlnAspGlnArgThrGln-----GlnProAlaAspSerAlaGlyGlnValGln 671
QY 2296 CAATCTTTAAAGAGATGCATACAGAGAA 2325
Db 672 GlnGlnGlnAspAspAspAspAspYrAspGln 681

```

Search completed: April 28, 2003, 11:29:39
 Job time : 1190 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4-P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 34.0331 Seconds

(without alignments)
4316.649 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407

Sequence: 1 MGSCRYFIIRMGNAASLE.....IDLINDSLVILNLLIYMKY 3542

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

112892

al number of hits satisfying chosen parameters:

0

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	617	3.2	1435	1	EBAL_PLAFC
2	435	2.2	2869	2	RBP1_PLAFC
3	415	2.1	6669	1	NRBU_HUMAN
4	381	2.0	3135	1	S230_PLAFC
5	357	1.8	3210	1	CENF_HUMAN
6	342.5	1.8	1070	1	PYDR_PLAFC
7	337.5	1.7	1153	1	PVDB_PLAFC
8	332.5	1.7	1875	1	MLP1_YEAST
9	327	1.7	1070	1	PVDB_PLAFC
10	326	1.7	1073	1	PVDB_PLAFC
11	317	1.6	1391	1	MST2_DROH
12	313.5	1.6	2116	1	MYSD_DROH
13	309.5	1.6	2748	1	NM1_YEAST
14	309	1.6	2704	1	BPAL_HUMAN
15	299	1.5	2492	1	ATRX_HUMAN
16	298.5	1.5	2663	1	CENF_HUMAN
17	289	1.5	1630	1	MSPI_PLAFC
18	289	1.5	1630	1	MSPI_PLAFC
19	285.5	1.5	1957	1	YD86_SCHPO
20	284.5	1.5	2476	1	ATRX_MOUSE
21	274	1.4	2230	1	GOG4_HUMAN
22	271.5	1.4	1790	1	USO1_YEAST
23	270.5	1.4	3685	1	DMD_HUMAN
24	269.5	1.4	2022	1	ANT1_ONCO
25	267	1.4	3678	1	DMD_MOUSE
26	261	1.3	1956	1	ATX1_PLAFC
27	259	1.3	3660	1	DMD_CHICK
28	259	1.3	5430	1	ACF7_HUMAN
29	258.5	1.3	1638	1	YK67_YEAST
30	258	1.3	1679	1	YK67_YEAST
31	258	1.3	2871	1	DESP_HUMAN
32	256.5	1.3	1727	1	ALM1_SCHPO
33	256	1.3	1805	1	HMW2_MYCBE

34	255.5	1.3	1557	1	DVAL_DICVI
35	254	1.3	3680	1	DMD_CANFA
36	252.5	1.3	1928	1	MYSD_YEAST
37	252	1.3	3418	1	BR2_HUMAN
38	252	1.3	3911	1	AKR9_HUMAN
39	249	1.3	5327	1	ACF7_MOUSE
40	245.5	1.3	1726	1	MSPI_PLAFC
41	243.5	1.3	2198	1	YK67_YEAST
42	243.5	1.3	2339	1	RPL2_CABEL
43	243	1.3	2349	1	TPR_HUMAN
44	242.5	1.2	1726	1	MSPI_PLAFC
45	237.5	1.2	1701	1	MSPI_PLAFC

ALIGNMENTS

Query Match	Score	DB 1: Length	DB 2: Length	DB 3: Length
1101	1101	1031	1031	1031
20	1031	1031	1031	1031
1158	1158	1158	1158	1158
63	63	63	63	63
1213	1213	1213	1213	1213
97	97	97	97	97
1272	1272	1272	1272	1272
150	150	150	150	150
1332	1332	1332	1332	1332

```

Db 202 EYMKDHEIFASKE---SOLLKKNNDKNYS-----KFNNDKNSFLDYGHLA 246
QY 1392 FGTDSKGGEGSKLEQIDSLFKNDOKSPNGKT---ROEWWTEBSHEIWEAMICALVK 1448
Db 247 MGDNDMDFG-GYSTKAMENKIQEYFKAGHGEISEHKIKNFKEWNNERREKLMZML----- 300
QY 1449 IGAKKDDFLENTGYNNVKSDDKSTLLEPAKRPQLRWLTLEWYDYCYTRYKLYADVOEK 1508
Db 301 -----SHKKNIN-NCKNIPDEEL-----QITQWIKENHGEFLERDRNSKLPKSK 345
QY 1509 CKNNDOLK-CDEECNKKCEDYKY-MKKKEWIPODKYKDKRDRKRRDRONIGWAVDY 1566
Db 346 CANNLTYEACEKECIDPCMKYKRWMTIRKFEHHTLSKETETOKVKE----- 392
QY 1567 TGTNADTYLNKFTASCCKGPGSASVYORNIOLLEKQAYYADKKQCC---TKFTE----- 1619
Db 393 ---NAENYL-IKISEKNND-----AKVSLLLNNDCAEYSKYCDCKHHTTYLAKSVLN 439
QY 1620 -ND---DKYTNISSKDKCK-GLVKEA-WTGAIKQNKGNPNYNNIKELTEDYLFPSRRL 1672
Db 440 GMDNITKEKREHIDDDSKFGCDNSVDTNKTWECKNPIYS-----TKDYCVPPRRQ 494
QY 1673 RICEFALDGNVTDPEVKDENG-L-RKRLMEVATEGYNLGYKKKEKKEKIKTSDAHKY 1730
Db 495 ELCLGNDRIRY-----DKNLMIKEHILAIYESRIKRYKKNKDKKE----- 538
QY 1731 STEVPPCSAMKSYDLNDIIGIDNLEDEKQTEENLKIFPNKNGSVGKSGDSTGNP 1790
Db 539 -----VCKIINTFADIDIDIIIGTGYMND---LSNRKLYVKINTSKYVHRKKN----- 585
QY 1791 GSTARFPMNNEKECVNNAKMGYKRGDDGNSASDEDLKKGSPVSDDYPMKGN 1850
Db 586 DKLFEDEMKKIKYKQVNVNVSVEF-----DKTYVKE 617
QY 1851 RDEGTAFOFLRFAMGEDEFCNKEKELEKLYGACNDYTCGDNEDKRRKCTDACYQYKF 1910
Db 618 DDIENIPEFERFESMGDYCODKTKMETLKEVEKERCEDD-----NCKSKSYKEM 672
QY 1911 ISEKMPQYKQKKGGE---NKDKIYSEHPYAKAEAREYLDKOLKICENNSGDCY 1966
Db 673 ISKKEEYNKQAKQOYEOKQNGNKKMYSEFKSIKEVYLYKIKYSEK-----CSNLFEDF 727
QY 1967 K-----C-NKDYSTORLTGDSNOMPASIDDEPKFV---ECKKQNGYVRG 2007
Db 728 KEELHSDYKNNCTMCPBKVDPIIS-IINNNQTSQEAIVPEENTELAHNTEPPSISEGPKG 786
QY 2008 PPRVRRPSPRVLISATASKEAKTAPPTKQPKVENLTTEMRQOTRRRAAOQTRK 2067
Db 787 NEOKERDDDS-----LSKISYS-----PENSRRP-----ETDAK 814
QY 2068 RTSATTTESVVG-TMVKAILISNKPDSKNGLEGCPKTYGQYPMWGCIVGSKENENGIC 2126
Db 815 DTSMLKIKGVADISMPAAVIGSSPN-----DNIN-----VEOGNIGISV- 855
QY 2127 MPPRRKILCINNIOYLYETENKRDNDIKFAIKCAIETQFLMKYIIEPAENELON 2186
Db 856 -----NSRPLSDVAPDKKELEDONSDESEYV-----NHSISPSINNDDGS 899
QY 2187 GTIDDEFKRLMYTYGYKDMFFGTDISNDKILTYVNSVTIINENNNKKQDKKDEEL 2246
Db 900 GS-----GSATVSESSSNTGLSIDDRNGDTPVROTANTAEVYIREKNMDKDE 951
QY 2247 RKIRWEKNKFIWGMITYGLYHILTDEN-----EKEKIRDNQYDMWTKLAPSLPEFYKRP 2302
Db 952 KGADEER-----HSTSELSPEKEMLTDEGNSLN-----HEEYKEH 990
QY 2303 QFLRMTEMAEFCNKRREOLK-----LEAGKEVEGNSN----- 2339
Db 991 TSNNDNVQOSGIVNMMVEKIKLTLLENPSSSIDEGAHHELSEPNLSSQDMSNTGCP 1050
QY 2340 DKGTOEACACVYTONFIKKWKTETEROEREFKDD-----KDGKKYKDYPSSTER 2388

```

```

Db 1051 DNTSEETER--ISNNEXKVEREDEFITKEVEDIVLKSNNRESDDGELYDE-----NS 1104
QY 2389 DIEKATCAHEYINMKTKELCGKOCSCMOKPSOLPKTTOOSOSDANDPESID-----Y 2444
Db 1105 DLSTVNDSEDAEAKMK---GN-DTSEKSHNSOIESDOCK-----NDMKTVDGLGTH 1155
QY 2445 VPEEFNKCEPELSK-----GSMHTKTKTEPKIPNCKEAAAYLSKEAENNDITLK 2499
Db 1156 VONEISVPTGEIDKLEKESKESKH--KAEERLSHDHK-----INPEDRNSWTMLK 1209
QY 2500 EKFTIESTEKESKNSMTNN 2521
Db 1210 D-----IRNEERHILN 1224

RESULT 2
RBP1_PLVAB
ID RBP1_PLVAB STANDARD: PRT; 2869 AA.
AC Q00798;
DT 01-Apr-1993 (Rel. 25, Created)
DT 01-Apr-1993 (Rel. 25, Last sequence update)
DT 01-Oct-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBP1.
OS Plasmodium vivax (strain Belem).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX NCBI_TaxID=31273;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9231538; PubMed=1617731;
RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
RT "A reticulocyte-binding protein complex of Plasmodium vivax
RL Cell 69:1213-1226(1992).
CC CC
CC CC
CC CC
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.slb.ch).
CC CC
DR EMBL; M88097; AAA29743.1;
KW Malaria; Receptor; signal; Transmembrane.
FT SIGNAL 1 17
FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
FT DOMAIN 18 2807 EXTRACELLULAR.
FT TRANSMEM 2808 2826 POTENTIAL.
FT DOMAIN 2827 2869 CYTOPLASMIC.
FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCEP CRC64;

Query Match 2.2%; Score 435; DB 1; Length 2869;
Best local similarity 18.0%; Pred. No. 2, 2e-12; Indels 984; Gaps 129;
Matches 529; Conservative 454; Mismatches 978;

QY 14 MAASLEGDAKSPLIKESHKSARNVLERYAKNIRH-PSKYA-KEHVDLSKGLDTKAEFRG 71
Db 428 DAAKKLES-----INDEBSAEFIDQKFPDNKALANNVCIFQYIKTNEPIKRAY----- 478
QY 72 GPSTPVAKHNYTYPCNDLKHETMLRIDDVNLRHPCHEGREQNRDEDESECGKRIAN 131
Db 479 -BSKYTKSNELLSTIIDTIGKGTAL-----GQSTPDQ---EC-NKIKT 518
QY 132 --YKRRNDATAACAPPRRRHCDKN-----LEALDINTQNHDLGNVLYTAKREGS 182
Db 519 EAEKVDAA-----EDICEKNEQIYIEIPSEDEJTIDDKINDLQDLIDQKREYKDE- 569

```


QY 183 IYNNHFKGTSACALANSFADIGIVRGIDMFKPNVDKVTGLREVFKKIHDMGME 242
 DB 570 IYNNSEF-----ISNRKNIYE-----NLKETYEETELINDICK 601
 QY 243 YKNDYDPDGSNGYKRLREAMNVRNKRKWEAT--TCDASKS----- 282
 DB 602 LEND-----TSKVNYLMOIKIRINTEKRIDESLOTVKFEKYLIDSKREKIELKEFEKSV 658
 QY 283 ---GYPMOESNTPLESNPKCGHKQ-----GKPTNLDYVPOYLR-----WPDWGE 326
 DB 659 TEINRQODESARDL-----HEBQIKELIDMAKKAHYIKELLSLKGRSSYFFEMNE 711
 QY 327 ---EFCRKNIKLKKVDCSRNDRKRLYCSHNGDCTTTW-----KKGLIHL- 371
 DB 712 LINTASYDMWEGFSAKKER--ADNDINALYNSYREDINALIEVEKFEVTEKNESTLEML 769
 QY 372 -----DNKCTDSTCKV----- 384
 DB 770 KDBEMEEKLODAKETPAKLNFEVSDKLTDTYTKMSAEVTNAGIKKELOKOPENVHKKM 829
 QY 385 -----FEVWLGNOO-----EAFKROKEKEKEIOSYLSNDNKEVNNINSEYKQ 428
 DB 830 KERSDAFSTKFEALQNSMOQYNEGDAIEKHQNRSEKEEYFKNSVEEDLSRETEERQ 889
 QY 429 FYEKLK-----ETQ---YATNDFPLNL---NEGKYCKGL 458
 DB 890 EYFKHKNFSRRKGEISAEITNMREVINKIESOLANTYGVYKFEYSLIDQNEVSTAK--- 946
 QY 459 PGKBDITFNNSADDKGIEFVREY-----CQVCPGCVGDCIKYTHKSDND--- 504
 DB 947 -ALKEIVSOSLRDKIDQIETEFKETSVAVENTVSTIOSLSKALIDSLKRLNGSINNCKY 1005
 QY 505 -----RRVNNEDYK--PFGVKEPTNITVLYSGNEQDITQKLEPNFNSST 548
 DB 1006 NNDIDILRSKIKTLREYQKEMPKRGDKGENTALLKSLRDKMKRINKELNDGLNLSL 1065
 QY 549 NYKDKANNQWECY-----KDBN-----INCKLEONTTEINNDNPKT--ISHNEF 552
 DB 1066 DPKREDELK---FYSESKSIHLSDOKGPODPLNR--IDEMEDIKRDELAVNVOVIS 1120
 QY 593 ELWVYTLNLDITLKMNDKLTCTINNTTHCIDECNRNCLCFDRVYKQKEEWMNSIKKLFK 652
 DB 1121 ENKVTLEFKNNSVYIEAMSHI--NVAHGITS--NKNEIL--KSKAEVEDLNV- 1171
 QY 653 KKNIOOSYYSNINN-----LEBGFYFVMDKLDKD-----BAKMKELMENIKRKK 697
 DB 1172 ---QHEDYKVKVKNPENKQLEAIRGSMKLEKVIKHYSEMTOLESTANTLSNMGKE 1227
 QY 698 NEFSNLENNRDYLENAIELLDHAKETATICKDNNTNEACETISHNATTNFCVPRGSTOP 757
 DB 1228 NEHDLEBLNK--TKGOMRDIYEKLIKIAELKEGTVNELKDANEKANR---VBP---EP 1278
 QY 758 TKRIKEIAQYFRKSAVEEARN-----RGLHKLKGAHGIYKRGRRKDFDNCRIMI--K 812
 DB 1279 ERRI--IGHVLEIYIEKDKAGVVEAMSLKRIEKLIOETS---DQSONEYVTTISITK 1333
 QY 813 HSNRNLFSGNPGCDKGTDGIGQTRFVVGTEWEDBEHMKRKHEDYIMPRRRHICTSNL 872
 DB 1334 HLENAGY-----EDVIKRNEDESIQREKAKSLETL 1365
 QY 873 EHLQTDHPLNGNIVDDYVNNISFLGDVLSAKYEANKIIRMYEKNKNNLKPKEVTPDKHQ 932
 DB 1366 DEKKKLVQOVNMLQSAIGNAGIS-----KELWELKVLIELLSTNY 1408
 QY 933 TTICRAIRYSFADIGIIRGRLMERNGDKVLOGLHLETFVGNIHRSLSKGGNDKYNDDA 992
 DB 1409 SSI---LEYKRRSSSEYFESOL--ANGEFYKAGE----- 1439
 QY 993 PKYTLKRENNMEAN--RAKYWEAMKCDIKYLDKSGHOSSTOSSYCGYSDHTPLDDYIPOKL 1051
 DB 1440 -----EKNASARLAEAEKLEKQIYKD-----LDYSDIDDKVY--- 1471

QY 1052 RMWTEWAEWYCKVQ--KKEYDLKEKCKECKDKDNQCGCTKESGTCTCTEACNEYNDI 1109
 DB 1472 -----KIEGJIKREILKMKESA----- 1487
 QY 1110 IGLMKRQWNLISDKYKLEHQQAQMSVNSGSIASSSTAKNHDNNTYIEFLSELYQONGKS 1169
 DB 1488 LTFWEE-----SEKFRQM-----CSSHMEAKGKKKIE-----YLRKNGDG 1524
 QY 1170 NKSJSD-----ESAVIGTNTTYENVAGYLDHGTGNFDDCQSONER 1209
 DB 1525 GKANITDSQMEVGNVSAEHAFAHTEVAGVDTKAFCSIVAVYKMNLFHESLMEKY 1584
 QY 1210 ---CDEKSDGKNEXYAFRDKPODDHAGCGKSGSPTRY-QITKTKKAEKDTBEKTYV 1265
 DB 1585 KYCERKNND--EAEKYSAKLKP--YDGRKARVSENERKISELKERAKYEKK--ESSOLN 1638
 QY 1266 DILKENDGKQVDECHPKRKNNGYPRMOCGINLVEDPRVCMPPROKLOVHLANDNEI 1325
 DB 1639 DV--STKSLQIDNCRQOLDS-----VLSNIGRV-----KONALQYFDSADKSM 1680
 QY 1326 KKLQ--SOVNLKBAFTKSAAEETFFSWYKYSKDGEGNELDK--ELKEGKIPAPFLNSMFY 1382
 DB 1681 KSVLPISLGAEKSLDKVYKAKE-----SYEKNIETVQNMESRLNVEGSL----- 1726
 QY 1383 TFGDYDFLEGTDISG---HGEKSLKQO-----IDSLFKNGDQSPNGKTRQEMWTE 1433
 DB 1727 -----TDIDKRITDIENDLLKMKRYEGGLQIKENADR---KSNFELVGS 1771
 QY 1434 HSHIEWEAMCALVYKIGAKKDDFT-----ENYG-----YNNVK-----FSGKSP 1472
 DB 1772 EYNALDPPSTIFIKLKYDWTGDKNTGYKMEIHEFTKSYNLTIEHLSNADYSV 1831
 QY 1473 TLEBFAKRPQFLMLTMYDYCYTRQ-----KYLDVQEKCKSN--- 1512
 DB 1832 TFE-----KQSLRELAKEKEEHLRRREEAIFLLNDIKKYESLKLKEMMKKYSAEYEGM 1887
 QY 1513 -----DOLKC---DTECNKCEDYVYKMKKKKEMIPDKTYKDEROK 1351
 DB 1888 KRDTSVSOLVODMKTYIBELKTLNDISECSSVLNNVVSIVKRYKE---SKHADYRDA 1943
 QY 1552 KRFDROHIGWVWDYGTNTDYLNRKFTASGDKDQSA-----SVQORNI 1597
 DB 1944 NSMTESNV-----TLANTFLSDEAKISSGMEFNAEKSNFKTDLEITSVJSNSN 1994
 QY 1598 QLEKQAYYDADKHCQCTKEIENDDKYTNISSKDKGKGLYEANTGAIKMKNRPNYNN 1657
 DB 1995 ELKK-----IEODSNDV--IQKRESQJLAKAD-----DIYN- 2026
 QY 1658 LKELTEDVLEPSSRLRICFHALDGNITDPEYKDBNGILRKLMEVYATEGYNLGOYKREK 1717
 DB 2027 -----VIKLNENEKLEEA-----KNKE 2045
 QY 1718 EKEKIKTSDAHKYSEYPPCSAMKYSFYPDRLDILGDNJ-----DEKOTTEBLK 1770
 DB 2046 EYVSEKREBLKLSQYEGIRCFENFHLDDTELEENLKNWYTIYRDKKSERESLOE 2105
 QY 1771 IFNKNGTSVGKSGDSTGNGSTARKFEWENKECVWNNAMICGYKRGDRDNGSNGSARSD 1830
 DB 2106 MENEMNTY---SNSITQLEGIVASA---GSKEDI-----EKLERSN 2141
 QY 1831 EDLK-----KGSVPS-----DDDYPMGKRNDEBTAVOFLWPAWGEDCKHKEKE 1877
 DB 2142 EEMRNISEKISTIDSKVIEKNSSTIDELYKIGKRCQ-----AHWISLISYANMK 2190
 QY 1878 LEKLVGACNDYTCGDNEDKRKKCTDACTOYKFRFISEMKPOYERQIKKYGKNNKRIYSEH 1937
 DB 2191 TSKKLIMIN-----KKEKTEKCVDIYKONSSSTIDYVETLKGFGYSKSLFSSASR-2241
 QY 1938 VAKDAB-----DAREYLDKQIKTICEKNSGDCXYKCMKDVSTORLTGNSQNMMA---SLD 1990
 DB 2242 IYONADYVSVFAKHESINAIYRDIKELLYLPHQNSDIS---IVEGVOYNNMLALYDKLN 2298
 QY 1991 DEKREYEGKNCQVPRPPRYVRETPSPRVSLSKATASKKEKAKTAPTPQPKVNTLT 2050

```

Db 2299 EERKEND-----ELYNISETKLKOEHSIDYK-----PMIEHKGMN---2337
QY 2051 EMRAQTRRAAQQTRKRTSTATTESDVGTWKALISNPDROGIE-----2098
Db 2338 -----ETNNKSLKKEKKIKSVNDHMSMEEMKKNCKLKTPEVONINNTYSIEAEVKT 2393
QY 2099 -----GCNPTYGQYKRWGCIYKGS-----KENENGICMPRRKRLCINN 2138
Db 2394 LEBIDRDYGNQYIEVHEKKQFSILIDRTVALMDIEIFEKKN-----YN 2439
QY 2139 IOYLVNTEKRRNDIKKEAFKCAIETQFLMKYIENPAENELONGTIDPEKRIAT 2198
Db 2440 LMEVNETIHRVNDYIEKTNKLVQAKTE--EQLLENTKQDDMLQNIETKRSIIEY 2496
QY 2199 YTYGDKYDMEFGTIDISNDKIIIVTNSVTI--LNNENKKKQDKKDELRKIFWEKKN 2255
Db 2497 FENVKKKKKSILNDLYEOEELKIGEHLEIKRNVETLSYSYIDCKMEMSKNLEKKS 2556
QY 2256 KFIWEGMIYGLVHLLDEN-EKEKINDYQYNDMTKLTLSLEFY-KRQPLKPTW-- 2311
Db 2557 KMANYSITELEREANEINRDAQIND-----DDTILNSVLEALIQKRGDMDAIFESQMSA 2611
QY 2312 -----AEECRKRRKQDLKLEAGCKEYECNGSNDKTOECACVYQNFIKKMT 2362
Db 2612 DRNPNEYKSAEKYMNANILIQLEVKLREI-----GOLVQDSHIL-----S 2654
QY 2363 ETEROREKRRKQDKGKYDYSTERDEIKATCAHEYLNMKLEKCGNCKSCMOPSSQ 2422
Db 2655 ENMSKKSALIEKERTALRTSENRRREE-----RAVOEKMNDMP--QSEITH 2704
QY 2423 LKPTQOOSQANDMPESLDYVPEFNKCEPELSKKSMTHTKITEPKIPANCVEKA 2482
Db 2705 SCSGSGEGKESQDEGLTHDAGDSTS-----SAKGA--HELEEEETTPM----- 2751
QY 2483 AYYSKEAENNDITL-----KEKPIESTKE--KESKNSWTNNPCDPKPYAP 2531
Db 2752 -----EEPEMN-DNTLLGVDTJRSDPEMDHTENTODGYTQDISNSDEADILNGKFNPN 2805
QY 2532 DKYTG 2536
Db 2806 VRYAG 2810

```

```

RN [3]
RP STRUCTURE BY NMR OF 6610-6669.
RX MEDLINE-98179559; PubMed-9514727;
RA Pollou A.S., Millevol S., Gaudel M., Kolmerer B., Pastore A.;
RT SH3 in muscles: solution structure of the SH3 domain from nebulin.;
RL J. Mol. Biol. 276:189-202(1998).
CC -1- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MAINTAINING
CC THE STRUCTURAL INTEGRITY OF SARCOMERES AND THE MEMBRANE SYSTEM
CC ASSOCIATED WITH THE MYOFIBRILS. BIND AND STABILIZE F-ACTIN.
CC -1- TISSUE SPECIFICITY: MUSCLE-SPECIFIC. LOCATED IN THE THIN FILAMENT
CC OF STRIATED MUSCLE.
CC -1- DISEASE: Defects in NEB are a cause of the autosomal recessive
CC form of nemaline myopathy (NEM2).
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -1- SIMILARITY: CONTAINS 178 NEBULIN REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83957; CAAS8788.1; ALT_SEQ.
DR EMBL; M19668; AAA59916.1; ALT_SEQ.
DR EMBL; M19669; AAA59917.1; ALT_SEQ.
DR PIR; A29979; A29979.
DR PIR; B29979; B29979.
DR PDB; 1ARK; 28-JAN-98.
DR PDB; 1NEB; 24-DEC-97.
DR Genew; HGNC:7720; NEB.
DR MIM; 161650; -.
DR MIM; 256030; -.
DR InterPro; IPR000900; Nebulin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR Pfam; PF00880; Nebulin; 146.
DR PRINTS; PR00510; NEBULIN.
DR PRINTS; PR00452; SH3DOMAIN.
DR SMART; SM00237; NEBU; 181.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.
FT REPEAT 113 107 NEBULIN 1.
FT REPEAT 113 143 NEBULIN 2.
FT REPEAT 148 178 NEBULIN 3.
FT REPEAT 183 213 NEBULIN 4.
FT REPEAT 218 248 NEBULIN 5.
FT REPEAT 253 283 NEBULIN 6.
FT REPEAT 289 318 NEBULIN 7.
FT REPEAT 324 354 NEBULIN 8.
FT REPEAT 363 393 NEBULIN 9.
FT REPEAT 398 428 NEBULIN 10.
FT REPEAT 434 464 NEBULIN 11.
FT REPEAT 502 532 NEBULIN 12.
FT REPEAT 537 567 NEBULIN 13.
FT REPEAT 573 603 NEBULIN 14.
FT REPEAT 611 641 NEBULIN 15.
FT REPEAT 681 711 NEBULIN 16.
FT REPEAT 749 779 NEBULIN 17.
FT REPEAT 784 814 NEBULIN 18.
FT REPEAT 820 850 NEBULIN 19.
FT REPEAT 858 888 NEBULIN 20.
FT REPEAT 893 923 NEBULIN 21.
FT REPEAT 924 954 NEBULIN 22.
FT REPEAT 959 990 NEBULIN 23.
FT REPEAT 993 1023 NEBULIN 24.
FT REPEAT 1028 1058 NEBULIN 25.
FT REPEAT 1064 1094 NEBULIN 26.
FT REPEAT 1102 1132 NEBULIN 27.
FT REPEAT 1137 1167 NEBULIN 28.
FT REPEAT 1168 1198 NEBULIN 29.

```

FT REPEAT 1204 1234 NEBULIN 30.
FT REPEAT 1237 1257 NEBULIN 31.
FT REPEAT 1272 1302 NEBULIN 32.
FT REPEAT 1308 1338 NEBULIN 33.
FT REPEAT 1346 1376 NEBULIN 34.
FT REPEAT 1381 1411 NEBULIN 35.
FT REPEAT 1412 1442 NEBULIN 36.
FT REPEAT 1448 1478 NEBULIN 37.
FT REPEAT 1481 1511 NEBULIN 38.
FT REPEAT 1516 1546 NEBULIN 39.
FT REPEAT 1552 1582 NEBULIN 40.
FT REPEAT 1590 1620 NEBULIN 41.
FT REPEAT 1625 1655 NEBULIN 42.
FT REPEAT 1656 1686 NEBULIN 43.
FT REPEAT 1692 1722 NEBULIN 44.
FT REPEAT 1725 1755 NEBULIN 45.
FT REPEAT 1760 1790 NEBULIN 46.
FT REPEAT 1796 1826 NEBULIN 47.
FT REPEAT 1834 1864 NEBULIN 48.
FT REPEAT 1869 1899 NEBULIN 49.
FT REPEAT 1900 1930 NEBULIN 50.
FT REPEAT 1936 1966 NEBULIN 51.
FT REPEAT 1969 1999 NEBULIN 52.
FT REPEAT 2004 2034 NEBULIN 53.
FT REPEAT 2040 2070 NEBULIN 54.
FT REPEAT 2078 2108 NEBULIN 55.
FT REPEAT 2113 2143 NEBULIN 56.
FT REPEAT 2144 2174 NEBULIN 57.
FT REPEAT 2180 2210 NEBULIN 58.
FT REPEAT 2213 2243 NEBULIN 59.
FT REPEAT 2248 2278 NEBULIN 60.
FT REPEAT 2284 2314 NEBULIN 61.
FT REPEAT 2322 2352 NEBULIN 62.
FT REPEAT 2357 2387 NEBULIN 63.
FT REPEAT 2388 2418 NEBULIN 64.
FT REPEAT 2423 2453 NEBULIN 65.
FT REPEAT 2456 2486 NEBULIN 66.
FT REPEAT 2491 2521 NEBULIN 67.
FT REPEAT 2527 2557 NEBULIN 68.
FT REPEAT 2565 2595 NEBULIN 69.
FT REPEAT 2600 2630 NEBULIN 70.
FT REPEAT 2631 2661 NEBULIN 71.
FT REPEAT 2666 2696 NEBULIN 72.
FT REPEAT 2699 2729 NEBULIN 73.
FT REPEAT 2734 2764 NEBULIN 74.
FT REPEAT 2770 2800 NEBULIN 75.
FT REPEAT 2808 2838 NEBULIN 76.
FT REPEAT 2843 2873 NEBULIN 77.
FT REPEAT 2874 2904 NEBULIN 78.
FT REPEAT 2909 2939 NEBULIN 79.
FT REPEAT 2942 2972 NEBULIN 80.
FT REPEAT 2977 3007 NEBULIN 81.
FT REPEAT 3013 3043 NEBULIN 82.
FT REPEAT 3051 3081 NEBULIN 83.
FT REPEAT 3086 3116 NEBULIN 84.
FT REPEAT 3117 3147 NEBULIN 85.
FT REPEAT 3152 3182 NEBULIN 86.
FT REPEAT 3185 3215 NEBULIN 87.
FT REPEAT 3220 3250 NEBULIN 88.
FT REPEAT 3256 3286 NEBULIN 89.
FT REPEAT 3294 3324 NEBULIN 90.
FT REPEAT 3329 3359 NEBULIN 91.
FT REPEAT 3360 3390 NEBULIN 92.
FT REPEAT 3395 3425 NEBULIN 93.
FT REPEAT 3428 3458 NEBULIN 94.
FT REPEAT 3463 3493 NEBULIN 95.
FT REPEAT 3499 3529 NEBULIN 96.
FT REPEAT 3537 3567 NEBULIN 97.
FT REPEAT 3572 3602 NEBULIN 98.
FT REPEAT 3603 3633 NEBULIN 99.
FT REPEAT 3638 3668 NEBULIN 100.
FT REPEAT 3671 3701 NEBULIN 101.
FT REPEAT 3706 3736 NEBULIN 102.

FT REPEAT 3742 3772 NEBULIN 103.
FT REPEAT 3780 3810 NEBULIN 104.
FT REPEAT 3815 3845 NEBULIN 105.
FT REPEAT 3846 3876 NEBULIN 106.
FT REPEAT 3914 3944 NEBULIN 107.
FT REPEAT 3949 3979 NEBULIN 108.
FT REPEAT 3984 4014 NEBULIN 109.
FT REPEAT 4021 4052 NEBULIN 110.
FT REPEAT 4057 4087 NEBULIN 111.
FT REPEAT 4088 4118 NEBULIN 112.
FT REPEAT 4123 4153 NEBULIN 113.
FT REPEAT 4156 4186 NEBULIN 114.
FT REPEAT 4191 4220 NEBULIN 115.
FT REPEAT 4226 4256 NEBULIN 116.
FT REPEAT 4264 4294 NEBULIN 117.
FT REPEAT 4299 4329 NEBULIN 118.
FT REPEAT 4330 4360 NEBULIN 119.
FT REPEAT 4365 4395 NEBULIN 120.
FT REPEAT 4400 4430 NEBULIN 121.
FT REPEAT 4435 4465 NEBULIN 122.
FT REPEAT 4471 4501 NEBULIN 123.
FT REPEAT 4544 4574 NEBULIN 124.
FT REPEAT 4575 4605 NEBULIN 125.
FT REPEAT 4610 4640 NEBULIN 126.
FT REPEAT 4645 4675 NEBULIN 127.
FT REPEAT 4680 4710 NEBULIN 128.
FT REPEAT 4716 4746 NEBULIN 129.

Query Match 2.1% Score 415; DB 1; Length 6669;
Best Local Similarity 17.7%; Pred. No. 5.4e-11;
Matches 695; Conservative 463; Mismatches 1299; Indels 1440; Gaps 180;

QY 265 VNKNKWEAITCDASYKSGYFMOSNTPLFSNPKCGHK---QGVPTMLDYVPOYLW 320
DB 66 VERKRYIRKKVDSKFTPTIASHQKQDLFSNKKYKEKTKGQYASTDTPELRR- 124
QY 321 FDMGGEFFCKRKRIKIKKAVDSCRNDKERLYCSHNGHDCITTTIWKGIHLDNKCDTDCST 380
DB 125 -----IKKVDOLSEVYRW---DGVANT-----ICHVDERKADIEH 159
QY 381 KCAVEFWLGNQOEAFFKKOKEKEKIQSYLSNDN-----KPVNN--INSEYKOPYEK 432
DB 160 AKRYV-----SQVSKYLYQNMEDTKOKYLLPPDAPELVQAVKNRMSKKLITDEWEA 213
QY 433 LKFTOYATNDT-----FLNLNKGKCYCKGGLPGEKDTITFTNSADKGIFFRSEYQV 484
DB 214 DKSLFYPYNDSPSLRVAQAKALSDVAYKKG--LAQQAQFPLPADP----- 259
QY 485 CPDCCGYKCDGKITKTHKSNDREXVNNEDYKPPMGVAKPTNITVLYSGNEGODITOKLENEFC 544
DB 260 -PD-----IEFAKRYTN---QVSKOKYK----- 278
QY 545 NSSTNYKDKNNOKWECYKQENINROCKLEQNTTEINNDPKIISPHNFEFLWYLLRDTI 604
DB 279 ----EDIEKIKKGM-----SETPCEYANANNDNDNISTKRYQODEF----- 317
QY 605 KANDKLKTKCINNNTTHCIDBCNRRNCICFDRVWVOKEBEWSISIKKLFKKKNIQOOSYYSNI 664
DB 318 --NMKQIYFMQRET-----PEPKMN-----KKAGVAAS----- 344
QY 665 NNLFEGYFFKVMKLDKDEKMKKELMENIRKK--NEFSMLENNRDYLENAIELLDHLK 722
DB 345 -----KVKYKEDYKKNKGKADYNVLPASENDQ-----LRQK 376
QY 723 ETATICKDNNTNACETSHNATNPCKVPRGQOPTFNKIEIAOVFRSAVEARNGH 782
DB 377 AADGALSDKIKYENETKAKKSINCEYK-----FKLDTVLQNF--SSDKYKDSYK 428
QY 783 KIKGKAHEGIIYKRGGRKDEKDNLCRIMIKHSNRNLFSGPDGKGTGIGTFRVVG 842
DB 429 DILG-----HYVGSFEDPYHSHCMKVTAQNSKN--YKAVEBEDRCKGFFPQT---ITQ 477
QY 843 EWE-VDEPHNRKDHEDYINPPRRRHICTSNLEHLQIDDHPLNGNIYDDLVNNSFLGDVLL 901

Db 478 EYEAIKLDCKDHITYKVHDPKTK-----FLOYVDSV---LLOAQVNSQLOSLN 526
 Qy 902 SAKYEANKI-----IMYKKNL-----KGREYTDPRHOTTICAIR 940
 Db 527 KAHHESEKFCCHPRPTPAITQIKHNAVNLSDMLYKODWEKSRKAKFDIVDAIPLLA 586
 Qy 941 YSPADIGDIIIRGBDLMERNDAVKLOGLHETVGNIIKSLKGNKYNDAPRYL----- 996
 Db 587 ANRKNISDVAKYKKDYKKNKGM-----GVLSIND- PKMLHSLK 625
 Qy 997 -----KLEMMEMANRAVW-----EAMKCDIKYLKDSGH----- 1027
 Db 626 VAKNOSDRLKENYEKTKAKSNMYCETPKYQDLOLKNSEAKYKDL-YKVDVIGHYVS 684
 Qy 1028 -----OSTOSSCYSDHTPLDDYIPOKLMMTEWMAWYKVKKEKEDJLK 1073
 Db 685 MEDPYTHCKKVAQAQNSDSYKREYEDKGCYFPOIT-----TOEYDAIK 730
 Qy 1074 EKEKECKDKDNGOCYKTESGTCT-----KC----- 1099
 Db 731-KLDQCKDHITYKVHDPKTKFTAVTDSVLLLOALNTKQLOSLNRYAKHGEHFKCHIPAD 789
 Qy 1100 -----TEACNEYNDILGKKEOW-----NISDKYKELHE 1129
 Db 790 APOFIOHRNATNLSDNYKODWEKSKAKFDIKYDAIPLAKANTKTSVMYKKDYE 849
 Qy 1130 QAO-----MSVNSG--TEASSTAKNHDNRVIEFSELAYOONGSKNSKSGT----- 1174
 Db 850 KSKGKMGALSTINDPKMLHSLKAKNOSDRE-----YKDOEKSTITYTAPLDMQ 901
 Qy 1175 -----SDSAVIGTNTYENGALHDTGNDDOQONFEDKSDGKDNKTYAFRDPDCH 1231
 Db 902 VYQAKSKQALASDVYKHL-----LH-----SYSY--PPD- 929
 Qy 1232 DGACGSGSKSPRYQIKTKKAEKDECKT-VNDILK-----ENGGKROVEDC 1280
 Db 930 -----SINVDLAKKAYALQSDVEYKADYNSMKKCGWYFPGSLEMEKAKRASDI 978
 Qy 1281 HPKNSNGPYDQCGNINIVEDPRVCMPR--ROKLCVHFLANDNET-----K 1326
 Db 979 LNEKTYROHPDYL-KFTSIEDAPITVOSKINQAOBSDIAYAKKEEITIHVNLPPDLPQ 1036
 Qy 1327 KLOQVNLKEAFIKSAAAEFFSWYKSKDEGNELEKELKEGIPAPILSMYTTGCD 1386
 Db 1037 FLOAKVNVN-----ISEMMYKADLKOLSKKGYDLRTDAIP--TRAANAARQA 1082
 Qy 1387 YRDFLEGTDISKGH-----EGSKLKE--QIDSLEKNGDOKSPNGKTRQEWTEHS 1435
 Db 1083 ASDVOYKDYERAKGKMGVFOSLQDDPKLVHYMNAVAKIOSDREYKDYKSKYNTPHD 1442
 Qy 1436 HEIWEAMLCALVKGAKK-DEFTENYGYNNVAKFSKSTLEBPAKROFLMLWTEWDDY 1494
 Db 1443 -----MNVYAAKKAQOVVSVNWK-----HSLH 1166
 Qy 1495 CYTRQOKYLKADVOEKCSNDOLCKTECNKCEDYVYKRRKKKEWIP-----ODKYKD 1547
 Db 1167 HYL---YLPDAMDELSKNMMQIOSD-NVYKEDYNNMK-KOIGMIPIGSLDYEVYKAGD 1221
 Qy 1548 ERDKRRDRQOIGVAVDYDTSTNATDYLNRKFTASCGDPPGASAVYORNTQ--LLEKQA 1604
 Db 1222 ALNKKYK-ROHPDYL-----KFT-SIVDSPYVQAKONTKQVSDILYKAK 1264
 Qy 1605 YUDADKCGGCKFTENDKYTNISKDKCGLYVEANTGAIKMKONKPNNTNMLKELTED 1664
 Db 1265 GEDV-KH-----KIT--MSPLPOFLQAKCNAVYS----- 1392
 Qy 1665 VLFPSRLRIGCFHALDGNVTDPEYKDENGELRKLMEVAATEGYNIQOYKKEKEKEKIKT 1724
 Db 1293 -----DYCYKR--DWMHILIRKGNVNLGDAIPITAAASRNILASDYKKEAYEKS- 1340
 Qy 1725 SDAHKYSYEVPCSAMKYSYDLDLITLIGDNLDEKQKEENIKKTFPNKNGYSVGKSD 1784
 Db 1341 -----GKHVGFSLQDDPKLVHYMNAVAKIOSDREYKKNYENTKSYHTPGD 1386
 Qy 1785 STTGNGSTARKFEMNENK-----CVWAMMIGYKRGGRDNGSGNSARD----- 1830
 Db 1387 MYTITPAKKAQOVATNVNKPOLHHYTYLPRDMSLEHRRNVQIOSDNYVDEYNSFLKG 1446
 Qy 1831 -----EDLKCSGVSDDDYPMKKNDEBTAVOFLMFAWEGEDFCNKEKELE 1879
 Db 1447 ICMPIGSLVEYKVRKADALMERKY-----ROHDPYKFTSVPSMOMMLAQHNTQOS 1501
 Qy 1880 KLVGACNDYTCGDNDKRRKCTDACTOYKKT-----SBMKRYEKOKIK-YGE 1927
 Db 1502 DL-----NYKVGEKLIKRYTID-PELPOTLOAKVNALNMSDAHYKADWKKTYRKGYDL 1554
 Qy 1928 MKDKYSEHPAKDAEDAREYLOKLKICENKSGCEYKCMKDVSTQRLDGSQNMPA 1987
 Db 1555 RPDAL-PIV-----AAKSNRIASDCYK-----EAYEKAGQOYF-L 1591
 Qy 1988 SLUDEPKRY-----EGKCONQVPRG--PRVYRRETPSPRSLISKATSKKEAKTAPPT 2039
 Db 1592 SLQDDPKLVHYMNAVAKIOSDREYKKGAYASKYHTPLDMVSV-----TAAKSOEVATNA 1647
 Qy 2040 KOPKVENLJTEMRA-----QTRRAAQOTRRKRTSTATTTESDVTM-VKALISNKPSR 2094
 Db 1648 NYRQSYHHYTYLPRDALNVEHNRNMQIOSDMLYKSDFTNMKKGIGWYPIESLEYEKAKKA 1707
 Qy 2095 GIEGCKNKTGYOY-----KMGCTVGSKRENE 2123
 Db 1708 GEI--LSKKKROHEKIKFTYAMDTEQALNSKNLMDKRYTERKW-----NKDKT 1758
 Qy 2124 GICPPRRKIKLINIOTL-----NYTEKRNDRINKENIKCALET----- 2166
 Db 1759 TIHWPDPIILSRVNGITMSDKLYKAGWEEKKGYDLRPAIATKAARASDIASDY 1818
 Qy 2167 -----QFMKTYIENPAENELONGTIPD-EFK----- 2194
 Db 1819 KYKRAYQAKGKHLGFSLEDDPKLVHFMQVAKQOSDREYKKGYSKTSFHTPYVDMLSY 1878
 Qy 2195 -----RIMYTY-----GDYKDKFPG----- 2210
 Db 1879 VAAKSOEVATNAVYRNVHTYNNLPDAMSEFLAKNMQIOSDNYKADADPMKIGWL 1938
 Qy 2211 -----TDISNDK-----KITVYNSVTIILNENKKNKODKKKDELRK 2248
 Db 1939 PLGSLAEKNNKAMEIISEKKYROHPDTLYSTLDSMNVNLAQNNAKIN--NEHLYK 1994
 Qy 2249 IFWE--KNKKFIMGW-----YGLTYHULTDENEKIND----- 2281
 Db 1995 QAMEADTKVHIMDIPQIILANAMALINISDKLYKLSLESKKGYYDLRPAIPIRAKA 2054
 Qy 2282 -----NYQVN-----DWTKLPISLEFVKRPOFLRMFTEMAEEFCUK 2318
 Db 2055 SRDIASDYKRYNVEKGGKMGVGFRLJEDDPKLVHSMQ-VAKMOSDR--EYKKNYENT 2109
 Qy 2319 RKE-----OLKLEAGCKEYECNSDGTQCEACAVTYONEIKRW-----KTEYER 2366
 Db 2110 KTSYHTPADMLSYTA-AKDAQANITN-----TYVKHLIHXYIILLPAMNIELTR 2157
 Qy 2367 OREKFKDKDKRK-YKYD-----PSTERDIEKATCAHEYLMLKELCKGNKDCSCMOK 2418
 Db 2158 NMRRIOSDNEYKODIYEMWKGLGWSFAGSLEYEKAKKAEYAS-----DDKYROH 2207
 Qy 2419 PSS-OLPKTYQOSO-----SSDANDPESIDYVPEEFNKCECELKSGKSMITTKITTEK 2473
 Db 2208 PSNFOEKKLSDSDVIALAKONAHITNKHLYTL--DMNK-----DKIKIHMPPTPOLQAK 2261
 Qy 2474 IPWNCYKAAVYLSKAEENMDITLKEF-IPLESKKEESKNSVTNNPCPKRPVAPD 2532
 Db 2262 QNOTLYSOKLYLKWBE-----ALKKGYDPIVDIASVOLAASDIASDYKQGYRKO 2315
 Qy 2533 --KYIGR-----NPCNENRFRVAYE-W-----KC----- 2557
 Db 2316 LGHHVGFSLQDDPKLVLSMNAVAKQOSREYKKGDEKKTSSPYVDMGLVYLAKKCOEL 2375

QY	2558	----	YKNSFYOEKRVVCPRRREJMCJLJNDEIKIRLDSNVLYLKMVRRJARAN	2612			
Db	2376	VSDVDYKX	-----YLHWOTCLPQONDVYQAKRYTELQSDENLYKSD	-----EMLRGIGWS	2425		
QY	2613	IKNFNSENGCMANPI	ICDTMYKTSFADLGDIVGTMDLIRIGYLPPEVKEIKLYV	-----FEYI	2669		
Db	2426	PLGLEMEKXKRAS	ELISEKKYR	-----QPPDRKKFSPIDADIY	2466		
QY	2670	YKGRYMNKGRNKYND	VQFRSMMWDAN	-----KDIKAMTCAPBDAFLFKGR	2720		
Db	2467	LATMANNRSDRLYRA	-----WDKDTQIHPDPRDYLKANKLINTSDKLYM	-----	2517		
QY	2721	MDGEFRTLLIODG	CKHDDPVDVYIPOFRFMWTE	-----WSEYSCALMELEKEFKSCCH	2777		
Db	2518	--GEEL	-----KRGY	-----DLYPDALPIKAKASRILABEYK	-----BEFRQDGHNI	2563	
QY	2778	-----	CKTSDRCKNDYDENCKE	-----OCTRCO	-----E	2802	
Db	2664	GARNIEDDPKMMWS	MHVAKIOSDREYKKDFEKKTKTFSSPYMDLGVULAYKQTLVSDVD	-----	2623		
QY	2803	YKNVLYKMSL	-----FDIOS	-----NKYK	-----ELYEQPIYTK	2632	
Db	2624	YKNTLHWOTCLPQ	OSVDYHARQATDLOSUNLYKSDOLMKIGIMKTSOSLEDEKKNKATQ	-----	2683		
QY	2833	ISYDVQNFVOKL	TFKSECVSEFSEYELNITSK	-----CLNYK	-----FVENDGSSNIRTAYEET	2890	
Db	2684	ILS	-----DHY	-----YRQPDQFKFSSLMDSIPVYLAKNNAITNMHRLYTEAMD	-----KDTYVHIMDT	2740	
QY	2891	PKSYKEACSTLE	SKRPDLNCPDNDKQDCKELOFTGCSKNDYNDNLMDNNAIYLVN	SS	2950		
Db	2741	PE	-----VLLAKONKYNSEKLYKLEL	-----AKRGYDMRD	-----	2775	
QY	2951	DONKGVLP	IPRRRHLCSTRPYAVNYKKGDEIL	-----KKKVLTSFSGQLG	2999		
Db	2776	-----	ALPIKAKASNDIASSEKYKEGYRKQDGHNIIGARAIRODPKMMWSMHVAKIOSD	-----	2829		
QY	3000	OKYSEBELCFE	AMKYSYADSYDILKGTDMMDTSLSEKIKKIFETSNEATENKRTKTMENN	3059			
Db	2830	REYKRD	-----FEKMKKRF	-----SPVMDLVULAKQOTLVSDVD	-----YKNY	2870	
QY	3060	ROFIWNAHLG	YKATISKVULDEGCOLPKDETNQFLRWL	-----IEMAKQACEKKHVS	3117		
Db	2871	LHQ	WOTCLPQOSVDYHARQATD	-----LOSMDKYSDLQMHMGIGWSIGSLDVE	-----	2919	
QY	3118	LKTKCP	SPNSDNFEASELLROPCQONDIRKYTSL	-----NILIKNTMENLIK	-----YKOLKD	3171	
Db	2920	-----	KCRATE	-----ILSDPKYRQ	-----PDRKFTSVTDSLEQVILAKNNALNMKRLYTEAMD	2971	
QY	3172	QSSONIDNK	SEENVQSYIKSKNSQCALBLNDINELYGTCKNNENNEKEVLLAKLYPOLY	3231			
Db	2972	KDKTQI	HIMP	-----TPPELMAROKN	-----IYSELYTK	3003	
QY	3232	FVEDETHKN	NVLDGNIKKEEOYVRPALYFTPHVDVSFYQALFSTHVAQDPRKNDILK	3291			
Db	3004	ANBEAKK	-----	-----GYDLRSD	-----	3018	
QY	3292	SSISVYV	SALGLIALHEMKKKKSSVDLRLINLPQGYTCMPLETESKNRITPIRSQPKY	3351			
Db	3019	-----	ALPIVA	-----KABRDVI	-----SDY	3042	
QY	3352	GKTYIV	MEGTSGDEKDYMDLSSDITSSSESEYELDINDIYVGSRYKT	-----L	3404		
Db	3043	KQLEHNI	GANNIEDPRKMMWSMHVAKI	-----QSDREYK	-----DFEKMTKTFSSPYMD	3092	
QY	3405	IEVVL	ESKRDPIDSDTPSRNDPRTNRJTDDEM	-----NELKDFVSQULYPTENPNNY	3458		
Db	3093	LGVVLAKCQ	OTVSDVYKNTL	-----HEWTC	-----LPQNDVIH	-----ARQAVDLOSNDIY	3141
QY	3459	K	-----SADIPMT	-----EENTLYXN	-----PEKFIITSI	-----HROLIYKEISTY	3499
Db	3142	KSDOLM	RLGIGWPIGMDVAVCKRAAEITLSDNYIKROPDKIKETSVND	-----	3190		

Query Match	Best Local Similarity	2.0%: Score 381; DB 1; Length 3135;
Matches 626; conservative 431; Mismatches 1153; Indels 1256; Gaps 169;	18.1%; Pred. No. 7,7e-10;	
524	ITVLYSGNEQDITQLEKFCN--SSPNKKDKNNOKWECYKEDENINCKLEQNTFIND	581

[illegible]

D	805	FEMIPYKKEIKNKKENKKSIGNLVNNSVYYNKKEMKAKFFNYQYVH - PRSYDOLINTLCS	863
Q	1570	-----NADYTL-----NKKFTASCGDRFGSASVVOIRNIOLEKOAYTDADKHCC	1614
D	864	ILKKEESNLSISTSYLAVYVSIJNELNLSLEDFEESFVPIKKTIOVAOKNV - NNEEHYT	921
Q	1645	TKFEINDOKRYNISXKR-----CKGLYKELANTAIKWKONGPANNYNKLELTDLPFSR	1670
D	922	CDFFDKDKDYIPSTANGKRLFLICKRHLKEDDTTLKC - VNKTOYPNIE-----IFPK	973
Q	1671	RLRLCFHALDGNVTDPYKDENGELKRRKMEV-----AATEGVNIGQYK--	1714
D	974	-----TLKCKEYLLKDLTDIOYQMSKEFFKFNTOAKNLNTPYLLFP	1017
Q	1715	-----EKKEKEKITSDAHKSYEPPOSAAKYSFVYLDRIIIGINDLEB-----EKO	1764
D	1018	FNHIGKKELKNNPYKHK-----DYKFEQSSVLSPLSADSLKLLNFPDLO	1066
Q	1763	KTEENLKIKFIKNGTSVSGSDSTTGNPGSTARFFMWNENKCECVNMAJCGYKRGDRGN	1822
D	1067	EVVCTEKIRYLNLSINELSDNNTFSVTRQVPPYI - DIKEPFYPMFCGNNGK - BGN	1122
Q	1823	SG-----NSARDEDLKKGCSYPSDDOYPMKANDEGTAIQLFLWPMGEDPCKKEKEL	1878
D	1123	IGIVELLISKOEKIKGCNPHESLTD-----	1149
Q	1879	EKLIVACNDYCGNENKRRKCTPDCTAOYKKFISEWKPOYERKIKYGENKDKIYSEH -	1937
D	1150	-----FENISSD-----THECLH-----AYENDIIGF-----NCLETTHPN	1182
Q	1938	-VANDAEAREYLDKOLKICENKSGDCEYCKKADYSTORLTDG-----NSQNPASLD	1990
D	1183	EVEEVEDEAEIYLOPE-----NCFNN-----YKGLNSYDITILKMAQYININNKPTPELK	1235
Q	1991	DEP-----KEYVGKNGCVOPPGPPV-----RRETPS-----	2017
D	1236	IPPYLLEDEVTISQCTIKOYKIKYIITKNDTYLLKREVOSESTLDDKIKYCEHENFI	1295
Q	2018	-PRVSLI-----SKATASREAK-----TAPTRQD - RRV	2045
D	1296	NPRVAKTFDEVEVETCNKIKTEFNFN YIQIRPCKADGICYKNIQMYADYKIPTRVPOFKF	1355
Q	2046	EN-----LTERBATRTIRRAAQOTR-----RSTATT	2075
D	1356	NNEELHLKIPNSEMLHTKELILYNEEKVDLHFVPLPIYIKDIYEENIVCDMSKMW	1415
Q	2076	ESDY-GTWVKAJLSNKKRDSRGIEGC-----NPKTYGOYKMGCIYKSGKENENGICM	2127
D	1416	KNOLGKVIYIITYSKREOK - YKGSFDEHNMHSYKTIYVKNCCIIDAKKDLIGVC	1473
Q	2128	PPRRKTL-C-----INNIOYL-----NY-ETENKRDNDIKEAFI	2159
D	1474	PSGTLKLNCFKDAIVHTNLTINGILIKNLNLANFTYKHOFNYMEIPALMNDIS--F	1530
Q	2160	KCAIETQFLWKAIJINPAENMLO-----NCTIDEFRIIYUYGSD -	2203
D	1531	KCICVDLK--KKKYNVKSPLGPRVRLKYLKLIKIDNTVYGT - DONKYTL - TYMDLH	1584
Q	2204	-----YKDMFCTDISNDKIIYVNSVTYI-----LNENKKKQDKKDELRKIPWE	2252
D	1585	LSHKRYNLKELFH - DLGKKRPADPDANPSITIESINESNESGPFPGVDABHL--	1639
Q	2253	KNKPFIMGMIYGLTYLTDENKEKIRNDYONDMFKL - TFSLEEFYKRPQFLWPFTEW	2311
D	1640	-----ILEG--YDWMESLYDEOLEEYI-----YNDIESLELDIOYVLOVNL-----	1680
Q	2312	AEEFCNKKREOOLLKLEAGCKEYECGNSNDCKTQECACAVTYONFLKWKMTYERQOREKF	2371
D	1681	-----KAPKLMAEAOJHNNRHVCDSKN-----NLIVP-----ESL	1711
Q	2372	KKDKGKKYKYDPSYTERDIEKATCAHEYMLNMLKLELQGNK-----DCSCKQBPSSOL--P	2424
D	1712	KKK-----EELGGPNVHICVALLKPKPDLTLYVCP	1741

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
CC or send an email to license@isb-sdb.ch).

DR EMBL: 019769; AAA82889.1; -
DR EMBL: 030872; AAA82935.1; -
DR EMBL: 025725; AAA86889.1; -
DR Genew: HGNC:1857; CENPF.
DR MIM: 600236; -
DR InterPro: IPR001230; Prenyl_site.
KW Chromosomal protein; Nuclear protein; Centromere; Coiled coil;
KW Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1338
FT DOMAIN 1642 1746
FT DOMAIN 1862 2987
FT DOMAIN 2207 2368
FT REPEAT 2389 2568
FT REPEAT 3015 3032
FT CONFLICT 16 16
FT CONFLICT 250 250
FT CONFLICT 272 272
FT CONFLICT 611 611
FT CONFLICT 1494 1589
FT CONFLICT 1611 1611
FT CONFLICT 1811 1811
FT CONFLICT 2242 2343
FT CONFLICT 2335 2335
FT CONFLICT 2492 2492
FT CONFLICT 2545 2561

SEQUENCE 3210 AA; 367589 MW; 1108332496084334 CRC64;
(IN REF. 3).
ELNERYALHNDQEAEC -> SSMREMOPTMTKKPVS
(IN REF. 3).
NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
T -> A (IN REF. 2).
L -> Q (IN REF. 2).
G -> D (IN REF. 2).
MISSING (IN REF. 2).
MISSING (IN REF. 2).
V -> A (IN REF. 2).
V -> L (IN REF. 2).
ER -> DG (IN REF. 3).
L -> Q (IN REF. 3).
D -> N (IN REF. 2).
ELNERYALHNDQEAEC -> SSMREMOPTMTKKPVS
(IN REF. 3).

Query Match 1.8%; Score 357; DB 1; Length 3210;
Best Local Similarity 18.7%; Pred. No. 1e-08;
Matches 574; Conservative 481; Mismatches 1120; Indels 898; Gaps 151;

DB 605 KWDILKTCINNTTTCGIDECNRNCLCFDRWYKKEEEMNSIKLFTKKKNIQSYISNI 664
7 EKKESLPT-----RTLOKIQLELGO---LDKLKKEKQOQOFOLDSEAPQOTQVVE--- 56
665 NNLEGGYFFKVDKLDKDEAKKELMENIKRKKNEFSN---LENNRDYLE---NAIELL 717
57 NEKTEG-----TNLKRNRQRLMEICELEKTKRKISHLQVKESQVNFQSGQLNSGKKQ 110
718 LDHLKETATICKDN-NTNEACETSHNATNPQVPRG-GTQPTNIKIKELAOYFKRSAYEE 775
111 IEKLEQELKRRKSELEKSOQAASQADSVLPCNPQKIFETPLRP---SQYSSSKYED 166
776 ARNGLHLKRAHEGIIKRGRRKDFKDNLCRLMIKSHNNLGFSGNCPDQKGTGDIQ 835
167 -----LKEKYNKEVEE---KKRLAEVKAQAKAASGLT----- 197
836 TRFVVGTEWEDPEHMRKDEHDVIMPRRRICTSNLEHLOTDDHP--LNGNIYDILVNN 893
198 -----PQATNNHRDI-----ARHOASSVFSMOQETPSHLISNSQRTPIRR 239
894 SFLDVLISAKYEANKIIRMYKEKNLKGPEVNDPKHQIT-GRAIRYSF-----ADI 946
240 DF-----SASY-----FSGLEVTIPSSSTQIDGRDANSSTFGSSSPHL 279
947 GDITRGDLWERN--GDM-VKLGHELTVEGNHKSILKSGKNDKYNDAPKYLIRENMW 1003
280 LDOLKAONOELRNKINLELRLQGH-----EKEMKGVN-----KFOELQL 320
1004 EANAAYWEAM-----KCDIKILKDSGHQSTQSSYCGISDHTPLDDYITPQRLRMTE 1056

DB 321 QLEKAV-ELIEKELKYNLCRDELVATTQYDQASRTYA-----LEQKLKLTLE 369
1057 MAEWYKQVAKK-----YDKLEKCKECKDKDN-----GQCTKESGTGCTKCE 1101
DB 370 --DLSCQONMASARCSLQKIKERKEKEFQELSRQSRFQTLDEQICQMRARLQLELQ 427
1102 ACNEYV-----DIIGLMEKQNNIISDKYKELHEQAOQSVSGSLASSTAKNHDVNI 1155
DB 428 ARNMHVLADELDKLTSVKQOLENNLEBKQ-----KLCRAQARQASQIKENELRRME 482
1156 EFLSELQOQNGSKSNKSGTSDSAVIGTNTYENYGAYLHDGNPDQOSONEFCDEKSD 1215
DB 483 EMKKE--NNILKSHSEQARE--VCHLEALKNIKQCLN-----QSON-FAEENKA 528
1216 GDNKRYAFRDKRPOHDAQCGKSSKPRVQITKKAAB--EKDTCKTYNDILKEND 1272
DB 529 KNTSOTMLRDLOE-----KINQOENSLTEKIKLAVADLEKQDSCS--ODILKKRE 578
1273 GKQVEDCHPKKNSNGYPMQCGNINLVEDPRVCPPRRQKLCVHFLANDNEIKLOSQV 1332
DB 579 --HHLEQLNDKLSK-----KINQOENSLTEKIKLAVADLEKQDSCS--ODILKKRE 602
1333 NLKEAFIKSAAAEFTFSWYKSKDGEENELDKELKSGKIPPAFLRSMFTYFGYRDLFL 1392
DB 603 ELKR-----KEYE-----ELKEEK----- 616
1393 GTDISKGEGSKLKEQDISLFLKNGDQSPNGKTRQEWMTSHSHIWEAMCALVKGAK 1452
DB 617 -TLFCMSSENEKLTLOWSEKENSQKINHLCTLKQOIKSHYNNRVLTLEMD--R 672
1453 KDDFTENTGYNNVFKSDKSTLLEEFKRPQFLRMLEWYDYCYTRQKYLKDVQEKCSN 1512
DB 673 ENLSVEIRNLNNV-LDSKSVETEQKLAVMELQKAEESD-----QKHQKEINMCLKT 725
1513 DQKQDTECNKCEDYV-KYKKKKKEMLPQDKRYKD----- 1547
DB 726 SOL-----TSQVDELEHKIQLLSNEIMDKDCRYODIHAYESIRDLKSKDASIVTNE 779
1548 -ERDKRRPDRQ-----HICVMTDYGTNATDYLRKRKFTASGQDGPASVQORIOILE 1601
DB 780 HQRSILAFDQPRAMHSHANITIGQSGMPSRSCRLEAD--QSPKSAIILQNRVDSLEF 837
1602 -----KQATYADAKHC-----GCTFIENDDKTYN 1626
DB 838 SLESQKQNSLOKQCEELVQIKGEIEENILKAKQOMHQSFAVETSQRSIKLOEDTSAHQ 897
1627 I-----SKDKCKGLYVE-ANTGAIKQNKGPNNY--NNLKE--LLEDVLFPSRL 1672
898 VVAETLSALEKKEKELQDLNDKYETEOAEIOELKSNHLLLEDLSIKELQLLSETLSLEKE 957
1673 RICHALDGNVTDPEVNDENGLRRLMEVATGEGYNIQO-----YKKEKKEKIKITS 1726
DB 958 MSSIIISINKREIE-ELQOENGLKEINASLQKEMNLIQSESPANTIDREKSIISLSD 1016
1727 AHKSYEV--PCGAMKSYDYLD-----IILGIDNDEKQTEENLKR 1770
DB 1017 QYKQEKILILLQRCETGNAVEDLSQYKAAQKNSKLECLINCTSCENRRKMLQELKE 1076
1771 IFNKGTSVKGSDSTTGNPGSTARKFFMNEK--ECVNNAMICGKRRDQNSG 1824
DB 1077 AFAKEHQEF-----LTKLAFERRNONLMELETVOAL-----KEMTDNON 1119
1825 NSARSDBLKCGSVPSDDYIPGKRNDEGTAVOFLRWFAMWEDRCK--HKKELEKLY 1882
1120 NSKSEAGLQOEIMTLKEQNKQKEVND-----LLQENQELKVKYKTHQEQNLE 1170
1883 GACNDYTCGNEKRRKCTACQYKFKFISEMKPQYEQKQKTKKGENKDKTIYSEHPAKA 1942
DB 1171 SEPIRNSVKEERESERQOC-----NFKPQMDLEKVE-----ISLDS 1205
1943 EDAREYLDQKLKICEK-----SGDCEYKCKM-----DVSQRLTDGNSQNNPAPSL 1989

Db 1206 YNQLV---QLEAMLNKELKLOESKEKELCLHLOETIRGDLTSLNLDMSQ----- 1256
 QY 1990 DDEPKVEKNCQVP-----RGPRV--RRETPSPVSLISKATSK--KEATAPTRQ 2041
 Db 1257 -----EISGLKDEIDAEKEEYISGPHSLSTQDNNAHLQSLQTMKNLNELEKICEILO 1311
 QY 2042 PKVENLITTEM---RAQ--TRTRAAQORPKRSTATTESVDGVKALKSPKPSRGC 2096
 Db 1312 AEHYE--LVTELDNRSECEITATRKMAEEVGLKLNVEKILNDSDGLHGLVEDIP---GG 1367
 QY 2097 IEGCNPKTYGQYPMGCIYKSKENENGICMPRRKKLCINNIQ--YLANETENKRONDI 2154
 Db 1368 ERGEQONE--QHPVSLAPLDESNSYEH--LTLSDKEVQMHAELOKELSLQSEKILND- 1423
 QY 2155 KEAFICAALETQFLMKYIIPENPAEN-----ELON 2186
 Db 1424 -----OHQMSKMSLEQTVDSLKAENLVLTNLNFGQDVKEMQGLGELVPSLS 1478
 QY 2187 GTIPD-----EKRIMTYTGDYKDMFGCTDISNDKLIITYTN--SVTITLBNKK 2235
 Db 1479 SCVPDSSSLSLDSSFPYRALLBQTDGDM-----SLLSNLEGAVANQSCVDEFGSSLD 1532
 QY 2236 KRODKKDE-----ELRKLTFEKNKKF---IWEGMIVGLTH--LNDENEKIRONYOY 2285
 Db 1533 TYVDSLKAENLVLTNLNFGQDVKEMQGLGELVPSLSSCVPSDSSSLDSSFPY 1592
 QY 2286 -----NDMTKLTPSLLEEFVRKPOFLRMWTEAEEFCNRRKBDQLKLENGCKREYDNGS 2338
 Db 1593 RALLEQTDGMSLTS--NLEGVASANO-----CSYDEVFCSLQSEENLTRK-----ETPSA 1640
 QY 2339 NQDKTQECACACTYQNFIKKMKTEYERQERKKKDKKDYDSTENDIEKACAN-- 2397
 Db 1641 PAKGVELESLECVYRQSLKELEKMEQO--GIMKKEIQLEQLLSSER--QELDCLRK 1696
 QY 2398 EYL--NMKLKELGNCDCSCMQRPSSQLPKTQO-----QSSQSDANDMP-----ES 2441
 Db 1697 QYISENQMOQKLTSLVLEMSKLAKEKQTEBLSLEVARAQGLDLSLSLGDIDT 1756
 QY 2442 LDVYPERFKCEPBLSSKSGSMITHKITEPKIPMNCVEKAAYLSKEAEN--NMDITLK 2499
 Db 1757 EDALQGRNESC--DISKE--HTSETTE--RTPKHVHQIC--DKDAQDNLNDIEKI 1805
 QY 2500 EKTIPIESTFEKESKNS--WNNNPPCDPK-----PVA--PDKYIGRNPNC 2541
 Db 1806 TETGAVKPTEGSGEQSPDTNYPEDEKQSGSECSISLFSGPNALVMDLGNEDT 1865
 QY 2542 ENR-----EENRFKVD--XEMKCYKSKFYOE-----KRCVCPYR 2575
 Db 1866 HNLQLVKETSNNLNLHYIEDRKRVESLMEKELSKLHLOEQVMTKIEACI--- 1922
 QY 2576 REHMCILNIDEIKIRLKDSNYL-----LKVARTARNEGID-----IKNFN 2618
 Db 1923 -----ELEKIYVELKESNDSLSEKLEYPSCHQHOELQVETSEGINSLDEMHADKSSR 1975
 QY 2619 SENGCANPCLDPMKYSFADLGIVGCTDMLRIGGILPPEIKLYKVEFITYGKWRKKNK 2678
 Db 1976 EDIGDVAKANDWKKEFLDVEN-----ELSRSEKASIE-----HEMLTYE----- 2018
 QY 2679 GRNKYNDVOTFRSAMMDANRDIKAMATKAPK-----DAKLFKGMDOFEFRTLLQDK 2733
 Db 2019 --ADLEVVQI--EKLCELEKDNENKOKYIVCLBEELSIVTSRNOLRGDLMDMSKTTALDO 2075
 QY 2734 CGHKDDPVDYDIIPQFRMWTSEYCYKALMELEKFKKSCDH-----KTSDR 2783
 Db 2076 LSEK-----MKRQTOELSHQSECLHCLQVAAVEKTELL 2112
 QY 2784 CKNDYDNCKOCTKQCEKKNFLVKWLSLFDI-----QSNKYKLY---EQPIY 2830
 Db 2113 QTLSSDVSELKDKTHLOELQSLKDSQALSTKCELENOIALQNKELSLVKESSSLQ 2172
 QY 2831 TKISTYDHVONFOKLTFKFS--ECSVESFSEYLETSHKCLNFKFNENDGSSNRTAFEE 2889
 Db 2173 ARISESDY-----EKLNVSAALPAALVEKGFALRLSLTQEBVHQLRGIEKLRVRIEAD 2227

QY 2890 TPESYACSCSLPSSKNPLDNCPTDQNKDCKELQFTFCSKNDYDNNDNNAYLVLS 2949
 Db 2228 EKKOLHIA-----EKLKEREND-----SLDKRYENLER-----ELON 2261
 QY 2950 SDNKGAVLIPRRRHLCITRITAVNRYRKQKELKRLKLL--LTSAPSOGLGQKYSSEE 3007
 Db 2262 SEENOELVI-----LDAEN--SKAEVETLKQIEEMASLVFELDLVTLRSEK 2309
 QY 3008 LCFEAKKYSYADSDILKGTDMADTSLSEK---KIFETSNBATENRKTMENNRQI 3064
 Db 2310 NLTKQIOENOGQSLDKLSLFSKSLBEKEQAEIQKESKRAVE---MLONOLKELN 2365
 QY 3065 HAM--LCGYK--IATSKVTLDEGWCOLPKDETNOF-----LRWLE--MAKQAC--- 3108
 Db 2366 EAVAALCGDQEIWKATEQSID-----PRIEEOHLNRSIEKLRARLEADERKQOLCVLQ 2419
 QY 3109 -REKRVSDSLK-----TKPRSNEN--FEAS-----ELLRO--PGQNDIRK 3148
 Db 2420 LKSEHRAIDLKGRVENLERELIARTNOHAALAEANSKGEVETLAKIETGQSLRG- 2478
 QY 3149 ISLNLIL--IKNTMENL-----NIKYOLKQSSGNIDNKPSEENVOSYIKSDQCA 3198
 Db 2479 LEIDVYTISEKEDLNLNELOKQERISELEITINSFENILOKEQEKYQMKESSTAMEX 2538
 QY 3199 L--ELNDINEIVTGKNN-----ENNEKRY---LKK--LYPGLYVEDETHRNV- 3242
 Db 2539 LOTOLKEINERVALHNDQACRAKAEQNLSSQVECLELEKQALQGLD---DEAKNYIY 2594
 QY 3243 -----LDGNIKREE 3251
 Db 2595 LQSSVGLIOEVE 2607

RESULT 6
 PVDR PLAYS
 ID PVDR PLAYS STANDARD; PRT; 1070 AA.
 AC P22280;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Duffy receptor precursor (Erythrocyte binding protein).
 GN PVDR.
 OS Plasmodium vivax (strain Salvador I).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxId=126793;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91187056; PubMed=1849231;
 RA Fang X., Kaslow D.C., Adams J.H., Miller L.H.;
 RT "Cloning of the Plasmodium vivax Duffy receptor.";
 RL Mol. Biochem. Parasitol. 44:125-132(1991).
 CC 1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
 CC DETERMINANT.
 CC 1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC 1- SIMILARITY: HIGH, TO P.KNOWLEDGE DUFFY RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 DR EMBL; M61095; AAA63423.1;
 KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane.
 FT SIGNAL 1 20
 FT CHAIN 21 1070 DUFFY RECEPTOR.
 FT DOMAIN 21 1007 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1008 1025 POTENTIAL.
 FT DOMAIN 1026 1070 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 183 183 N-LINKED (GLCNAC...) (POTENTIAL).

2381	KDPSTEDIDKACCAHBYLMAKLE	-----LGCN	-----DSCMNKPSBQPK	2425		
79	K-YGTTELKITYAOMNGKRSSRIKES	IGYAHNFGNSYMEGKDGDKGEEKOEH	-K	135		
2426	TTQOOSDANDMESLDYVPEEFNKECEP	ELSKKSMITHKTEIPKIPMCVEKAY		2485		
136	TDSKTONGKANNLWMLDY	-----ETSSNQ	-----PAGTLDNVEF	172		
2486	LSKAEENMDTLLEKTEPIESTKEKESK	SWTNNP	-CDPKPYAP	-DKYIGRNP	2540	
173	VTGHEGNS	-----	RKSSNGSPYDIDHAKKISSAII	INHAFLOVY	213	
2541	CENREENFKYDYMCKYKSKSFYOEK	RRYCPREHDLNLTDEIKTERLNS	-----	2596		
214	MKNCKYKRRKRRERWDC	-----NKKYOCIPDRRYQJCMEL	-TWLVNNTDINHRD	264		
2597	-----YLLKMYRTTARNBSGIDII	KNFSENGCAMPICDTRKYSFADJIGY	GTMLR	2650		
265	ITFRKLTYKRLIYDAVEG	-DILLKLN	-NRYNKKDFCKDIRMSL	IGDFGIIIMTMEG	321	
2651	IGGLYPVEIKYKVFYIYGKWRNKN	KBNYNDYOFRSAMWANKKDIYKAMTCAP		2710		
322	I-GYSKYVENNLSIF	-----GTDE	-KAQORRKNWESKAQITWMTMTSV	365		
2711	EDAKLFRKGRADGERITLLIQDKCGH	KDPPVDYI	-----PQRFNTEPME	2758		
366	-----KKRLKG	-----	NFLWICKLVANVIEBQYIRHRI	IRMRG	398	
2759	YUCCALMELEKRRKSGCDHCKTS	SDRCKNDYDENC	-CE	-QCKTRCOEYKKNFVLKWSLFD	2815	
399	DYSELTEYEOUKLEKCGD	-----KIWYTRKCKKVPPOCMCK	SYDOWITTRKKNQMD	451		
2816	IOSNKYKELEOPIYKISYDHQNFVOK	LTFPSECSVSPFSYLHNSKCLNYKFN	E	2875		
452	VLSMKF	-----ISVKNNAEKQYTAGIYPIYLKOE	-LDEFNVAFEN	-EINK	496	
2876	NDGSSNIRIYAEETPKSYKEAC	SCTLPBKNPLDNCPTDOKDKC	KELOTFTECSKNDY	2935		
497	RDG	-----AYIELCVSV	-----EAKKNTQGEVY	521		
2936	NLNDNMNAYLVNLS	-----DDNKGVLIPRRRLCRPTTA	NYNRYKDKKEILKLLT	2989		
522	-NYDMAKSOATNSINISOPVSSA	KEKYPGDSTH	-----GNVSGDSSITGVAIV	572		
2990	SAFSOGOLLQOKYKSEBELCE	PFAMKYSYADVSDILIKGTDMDT	SLSEKIKIFETSNEAT	3049		
573	-----GDGQ	-NONGPAPBSDY	-----QSDIAESSAANVDPQKSVSK	609		
3050	ENKRTWMENNRQIWHAMLCG	KIKATSKYTLDEBGCQOLPKDET	NOFLRWLIEWAQACK	3109		
610	-----	-----RSDPTAS	-----VTGIEACK	625		
3110	EKKHVSDSLTKCPRENEDEFA	SELRLPOGCONDIRKYSINILIKNEMEN	-LNIKYO	3168		
626	ENIGASNS	-----RSESTEYAN	-SPG	-DDIYNASISIPV	-SEENLVTPYING	670

[illegible]

FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 626 626 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 847 847 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 856 856 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 910 910 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1153 AA; 130471 MW; 6497BD16CE7BEC01 CRC64;

Query Match 1.7%; Score 337.5; DB 1; Length 1153;
 Best Local Similarity 18.8%; Pred. No. 2.3e-08;
 Matches 237; Conservative 168; Mismatches 434; Indels 425; Gaps 53;

QY 2382 DYSTERDIEKATCAHEYLNMRKELCGNNDSCMGKPSQPLKPTQOQSSDANMPES 2441
 DB 40 EYKKNNGYKLAH-GHHYMD-----NDQIERMLQGTDRSRVAKTEEN 80
 2442 LDYPEEFNCEPELSEKSGSMITHKTEPKIPMNCVEKAAAYLSKEAEN----- 2493
 DB 81 VKY-----KYNVEELNTEKTEQTKGRIN-----RIKSTYEADQNVADNNTIDKANG 128
 QY 2494 -----MDITLKEKPIP-----TESTKE--KESKMSWTNNPCDPRKPY 2529
 DB 129 EYKTDKNTKNGEGARMMVMDYDISGSGHIDYVETGTEDEGFLNENSGKGDHPY 188
 QY 2530 APD-----KYGRNRCENREERFVDEMKCYKNSKFEYQSKRCVCP 2573
 DB 189 RMNRKERMCSGVINOTFLQKNVNR--CNKKRGRTR--DWDC-----PRKQVCIP 236
 QY 2574 PRREHMCRLNDEI-----KIERLDSNYLLKMYR---TARNEGIDIILKNF---SENG 2622
 DB 237 DRRAYQLCMKELTMYNNKTHSHNDITFLKLNKEKLTDAEAGDILLKKNYVSED- 295
 QY 2623 CANNPCLDITMKYFADLGDIVRGTDMIRIGYLPVYIKLYKFEYIYGKMRKNKGRK 2682
 DB 296 -----LCKDKIWSLEDFGIMGTDEGI-GYSQVVENMLRTYE-----GT 335
 QY 2683 YNDVOTERSAMPMANRRKIMKATCTKAPEDAKLFRRKGRMDGFERITLIDCKGHKDDPV 2742
 DB 336 GRTYQIDRKKMMWESKRYIWEATY-----LSVKKKLNGYSAM---NC--KEDVQI 380
 QY 2743 DDIYPOFRMTWMESEYCCALMELEKEKESCDH-----CKTSRCKNDYDENKC 2793
 DB 381 -NVEPQIYRIRRMGMDSYSELPEQQRKIKKCDRKLYTNLRLICMS----- 427
 2794 ECKTTCQCEYKKNVLMKSLFLDIQSNKYKELYQPIYTKISTYDHOVQNKLTREKSEC 2853
 DB 428 -PCNDSCKLDDWITRRKKQMDVLSYRFSVKKGOIT-----ETENITTAVIDIKOE- 478
 QY 2854 SVSFSFYHETSKCLNFKENENDSSNRTAFEEPRKSKAGSGTL--PSKNPLDNC 2911
 DB 479 -LNGFENWME-----NEINKKRDNV-----YIDICLCAADEPKMN- 512
 QY 2912 PTDOAKDCKELOTFTFGSKNDYDNLDMNNAVLVNSDDNKGVILPFRRLHLCRPT 2971
 DB 513 -----TQEHKTKLKS-----APKLETRGSHSTIQPMS 539
 QY 2972 AYNYRR--GDKELKKKLKLTSAFSGQLQKRSBELCFEAMK-----SYADYSD 3022
 DB 540 SSGAEVYVQGD-----LAHGNINDAAKSTTD--EAKGDDGONGNOTVASEN- 583
 QY 3023 IIRGTMDMDLSLSEKIKIPETSEATENRKYWENNRRQIWMALCGYKATSKVYLDE 3082
 DB 584 -IKGTDNIE-----NEAAKNVDT-----YKFTYERSADTR 612
 QY 3083 GWCQLPRDEETNOFLMLLEMAQACEKKHVSDESKTKCRPSNEDNFEASELLRPGCQ 3142
 DB 613 GADIDTETGE-----EKLNTSYSGSSE--ITYKEMIPGDIY 647
 QY 3143 NDIRKATISLNLKNTMENTNINIKY---QLAKDS-----SGNIDNKPSEEN-VQSYIKSK 3193

DB 648 KDV-----SAAVENENPLETKHKTIFEDSKDONSSENSGSEKFAATSNPTTEAVESS 701
 QY 3194 DSQCALELNDINIEIYGTGKNNENNEFEYVLRKLYPGLYFEDTTHNNHYLDG----- 3245
 DB 702 SAAGVOEDSAHNSVMTGRDNSTISATSDGLSSGDKRESLITSIEMADGGDPVQGSLL 761
 QY 3246 -----NIKEEQTVRPKALYFETPHVDSFYQAPLESTRHVAQYDPKN-D 3288
 DB 762 WNLNDPSVAGGKGKSHIKTEE-----NEGSAEIDDKNDV 797
 QY 3289 ILKSSISVYI-----YSALGLIALHFMKKKFKSSVDLRIILIPQGEFG 3332
 DB 798 IAOQRTATITTEVQPERPDLSDTNGVNPBSG-----NKQNGCAVALSCELSIESNEV 850
 QY 3333 MPLESKNRIIPRSGPYKQ--KTYIYMEGDTSGDE--DKYMMDDSSDITSESEYEY 3387
 DB 851 HKTIIDTTHGLEKKNKNGENKEDFOKHDPMMNDMLNDQSSDQISSDQISSNQTSSDQISSN 910
 QY 3388 LDINDIYVPSPPRYKTLIEVYLEPSKRDIPSDPTPSNDPRTNRFIDDENNELKHDFVSG 3447
 DB 911 QTSDD-----QTSDDQISSDQISSDQISSDQISSDQISSDQISSDQISSDQISSDQISSD 953
 QY 3448 YLEPTEPNNNYKASADIPMTEPTIYSDN--PEKFPITSIHRDLYTCKE--ISTYNIN 3502
 DB 954 RNFELKSSSEDMKSGDEMRNSNSNELYSHNNLNNRKLINDOYERHDVAKATREKILLMSEVA 1013
 QY 3503 MSTN 3506
 DB 1014 KCNN 1017

RESULT 8
 MBL1 YEAST
 ID MBL1 YEAST STANDARD; PRT: 1875 AA.
 AC 002455;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin-like protein MBL1.
 GN MBL1 OR YKR095W OR YKR415.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_Taxid=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=5288C;
 RX MEDLINE=93247549; PubMed=8483450;
 RA Koelling R., Nguyen T., Chen E.Y., Botstein D.;
 RT "A new yeast gene with a myosin-like heptad repeat structure.";
 RL Mol. Gen. Genet. 237:359-369(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94205265; PubMed=8154186;
 RA Bou G., Esteban P.F., Baladron V., Gonzalez G.A., Cantalejo J.G.,
 RA Remacha M., Jimenez A., del Rey F., Balleseta J.P.G., Revuelta J.L.;
 RT "The complete sequence of a 15,820 bp segment of Saccharomyces
 RT cerevisiae chromosome XI contains the UBI2 and MBL1 genes and three
 RT new open reading frames.";
 RL Yeast 9:1349-1354(1993).
 CC -1- FUNCTION: MYOSIN-LIKE PROTEIN THAT IS PROBABLY INVOLVED IN DNA
 CC REPAIR.
 CC -1- SIMILARITY: SOME, TO THE TPR ONCOGENE.
 CC -1- CAUTION: REF.2 MISQUOTES THE GENE NAME AS "MBL1".
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 DR EMBL: L01992; AAA34783.1; -
 DR EMBL: X73541; CA51948.1; -
 DR EMBL: Z28320; CA82174.1; -
 DR PIR: S38173; S38173.
 DR SGD: S0001803; M1P1.
 KM Coiled coil; DNA repair.
 FT DOMAIN 69 487 COILED COIL (POTENTIAL).
 FT DOMAIN 531 1678 COILED COIL (POTENTIAL).
 FT DOMAIN 1834 1866 COILED COIL (POTENTIAL).
 FT CONFLICT 301 301 R -> A (IN REF. 1).
 SQ SEQUENCE 1875 AA; 218455 MW; 683AD34C906867 CRC64;

Query Match 1.7%; Score 332.5; DB 1; Length 1875;
 Best Local Similarity 19.1%; Pred. No. 7.1e-08;
 Matches 445; Conservative 350; Mismatches 897; Indels 637; Gaps 100;

QY 1321 NDEIKKLGQVNLKEAFISAAAEFF--SWYKSKSGS-----EGNEIDKEL 1367
 DB 4 HDPPMESLQNGENSE---RLNALASFSGSLQVKSFGDGVVYKHLNDKLLQFNEKSE- 59
 1368 KEGKIPPAIRSMFYPPGYRDFLFGTDISKGHGSGSKLEQIDSLFKNGDQKSPNGKTR 1427
 DB 60 -NKVTVYVF-----DELKASSLKIDG---LKTENENYIRENDKIR---KER 99
 QY 1428 QEWMTSHSHHEWEMICA--LVKIGAKKDDFTNYGYNVKSDDKSTTLEFPAKRPQFLR 1485
 DB 100 NDFVVFESVENEMKMLISELEFVKRRKLDLTBEKKETOSNOORITKILDERLKEIEVR 159
 QY 1486 WLTEWYDYCYTQKYLKDVQEKCK--SNDOLKCDTECNKCEDVYKMMKKKEKTPQ- 1541
 DB 160 VENNRSNSECKKLRSTIMOLETKQGYTIND-LNSRTELEKTKOE--LTLLQSNNDLEKE 217
 QY 1542 ----DKY--YKDERDKKRPD--RQIRGVWVTDY---TGTNADYLNRRFTASCGDK-- 1586
 DB 218 LRSKNBOYLSTROKTDKVIDIRNELNRLNDFOMERTNNDVLQKNNELSLQEKLE 277
 QY 1587 ----PGASVVOYRNLTQLEKQAYYDADAHCCCTCFIENDDKYTINISKD 1631
 DB 278 IGLGSLNSEKQEPFAESLSKQRLVDLLESQ-----LNAVK 314
 QY 1632 KCKGLVKEANTAI--KWQNGPNNNYNNLKE--LTEDVLFPSRR--LRICFHALDGYT 1684
 DB 315 EELNSTRLEINTKAVIADSKKQTPENEDLTKELQTLKEKLAQCEKCEURL-----SST 368
 QY 1685 DPEVDEKNGLRKLEVAATEGYNL-GQYKKEKKEKEKIT--SDAKYSTEVPCCSAM 1740
 DB 369 DDAEDENELSKA---SSSDFTFLKQULIKERRKEHLQNOQIETFIYELEKRVPIINSF 424
 1741 KYSF---YDLRDIILGIDNLEDEKQTEENLKIFNNKNGTSVSGSDSTGNPSTARK 1796
 DB 425 KERTDULENELNNAALLLEHNSNEK---NAKYKELNNAKNOKLVEGENDLOT---LTKOR 477
 QY 1797 PFMMNNEKCVWAMAMCYKRRGHGDSGNSASASDEDLKKGSPVSDDDYPMGKNRDEGTA 1856
 DB 478 L-----DLCRQIQLYLLITNSVSNDSKGPLRKEE----- 505
 QY 1857 YQFLRFAMGHEDFCKHKEKELEKLVGACNDYTCGDNDKRRKCTGDACTQYKKTISEKP 1916
 DB 506 IQFIONIMQ--ED-----DSTTES--DSQKVVYTERLVERKKNIT----- 540
 QY 1917 QYENQIKRYGKNDKIYSEHPVADAEDAREYLDKQLKRICENKSGDCEYVCMQ-VSTQ 1975
 DB 541 ----OLQF-----KNAE-----LLKAVRNLAADKLESEKKSQSLQ 572
 QY 1976 RL---TDOGNSGNMPASLDDEKVEYKGCNCOVPRGPPRVRRRETPSPRVSLISKATASKE 2032
 DB 573 KIESETVNEAEAIITLKSEMDL-----SRIEELQKELEE 609
 QY 2033 AKTAPPTKQPKKVENLTETMAQOTRTBRAAO---OTRKRTSTATTTESDGVYKVAIIS 2088
 DB 610 LKTSYVP--NEDASYSNVITIKQLETFRKRLDESQVQDLQTRISQITRESTE----- 656

QY 2089 NKPDNRGIEGNCNPKTYGQYPMKCTIVGSKENENGICMPRRKRLCINNIOYLYNET-- 2146
 DB 657 -----NMSLNKEIOD 667
 QY 2147 --ENKRDNDIKFAFICAAI--ETQFLMKYIIEHPAENELONGTIDDEFRIWYVYG 2202
 DB 668 LVDSKSDISIKLGKESKSSILAEERKLLSNLIDLTKAEN-----DQLKRPDYL-- 717
 QY 2203 DYKDFEFGDISNKKIITVNSVTTLNENKRRKQDKKDELRKIFWEKKNKFIWGM 2262
 DB 718 -----QNTILKQDSKTHETLNEVYSCSKSISIVETE 748
 QY 2263 IYGLATYHLDENEKEKIDNVOYN---DNTKILTP--SLEEVKRPQLRW---ETEMA 2312
 DB 749 LNLN-----KEQKRLVHLEKLNKQELINKLSPKSDILRLWVQLOTLQKREDLLEET 801
 QY 2313 EEFCKRKKEQL-----LKLFAKCEY-----EENGSDNGKTOCABACVYQNFIKW 2360
 DB 802 RRSQCKKIDELDALSELKETSQGDHHIKQLEEDNNEN-----IETYNKIEAL 851
 QY 2361 KTEYE---RQREKFKKQDKKRYKDYPESTERDIEKATCAHEYLNNKIKELCGNRDCSCMQ 2417
 DB 852 KQDYESVITSVDSKQTDIEKLOYK--VKSLEKEIEEDKIRLHTYVNMDETI--NDSILKE 908
 QY 2418 KPSQLPKTQOOSDANDPESLDYVPEENKCECEPESKKSMTHTKITEPKIIPAN 2477
 DB 909 LEKSKINLTDAVSQKEKEDLYETTSOSLOOTNS-KLDE-SFKPTNQIKMLTDEKTSLE 966
 QY 2478 CVERAAVYLSKBAENNMDITLKEKFIPIESTREKESKSNMTNPNCPKRYAPDKYIGR 2537
 DB 967 --DKISLKEQMFNLNNELDQKGMEREKADFFKRISILQNN----- 1007
 QY 2538 RNPCEENRENFYDYDEKCYK--NSKPYQKRYKCVPPREHMLRLNDEIK-TERIKD 2594
 DB 1008 -----NKVEVAKSYTESKSLQIONDDQYTYANTOANNYEDLOKHABVSKTISELR 1062
 QY 2595 SNYLLKAVRTARNEGIDILNFNSEN-----GCAMPICDTMKSYFDLGIYVGTMLR 2650
 DB 1063 QLHTYKGGVQKTL-NLSRQLENALKENBKWSQKESLQDLQDLSNRIDLSQNKLLY 1121
 QY 2651 IGYLPPEIKLYVFEIYKMKRNKNGKRYKNDVOTFFSAMWANDARKDIWMAKTCAP 2710
 DB 1122 ----DQIDITYAAD--KEVNNSTNGPLNILLTLR-----REBDILTKYVAE 1165
 QY 2711 EDATLFRGRMDGPERITL-----IODKCGHKDDPPVD-----DYIIPQFRMT- 2754
 DB 1166 RDAKMLR-----QKISIMDVDELQDARTKLIDNRVAKENHSSIQOHDIKRLQNLNL 1218
 QY 2755 -----EMSEYCYALMELEKFKKSCDHCTSDRCXNDYDENKCEQKTRCOE 2802
 DB 1219 LRESNITLRLNELNNNKKELESELDKLQONVAPISEITLTKYSQOEKQELAKKE 1278
 QY 2803 YKNFVLKWSLSLFIQSKRYELVEQPIYTKISTYDHQNVQKLTFFKSCSVSESEYEL 2862
 DB 1279 ----VHRMKK-----RSODILKEHQ-----LSSSDE-----KLESEIENKEEL 1315
 QY 2863 HETSKC---LNYKFENEDGSSNIRTYAFEEETPKSYKACSCTLPSKNPDLNCPQDQNKOG 2919
 DB 1316 ENKROGAEAEEKFNRLRQAGRLKTSKLSQSLTEQVNSLDAKNVYLSNLSSEMA-R 1374
 QY 2920 CKELQTFPFSKDYDNNDLNNNAVYLVANSSDDNKGVLLIPRRRHLCSTRPITAYNRYKGD 2979
 DB 1375 IEBLQNAKVAQGN--NOLE--AIRKLEDAEKASRELOAKLEBSTSYESTINGLANE 1428
 QY 2980 KEILKKLLTSAFSOGOLLGQKYSSEEL--CREAKKYSYADYSIDIKGTDMADISLSK 3037
 DB 1429 ITLKEIEKQRIQOOLQTSANEQNDLSNYESKKSFEF--DKIKITKEKTOEVNKK 1486
 QY 3038 I-----KKIETSNEATENKRTWNNRQIWIWAMLCGYIATSKYVLIDEGWQOLPDEET 3093
 DB 1487 ILBAOERLQPSNINNEELKKWSE-----HBOEVSQKIRBAEELAKKR-IRLPTPEKI 1540
 QY 3094 NQFLRWLIEMAKQACEKKHYSDSLTKCRPSNEDNEFEASELLRQPGCONDIRKYSILNI 3153

DB 1541 NKTIERKELEKEFEERK-VEERIKS-----MEQSG-----ETDV 1574
 QY 3154 LKNTME-NINIKYKOLKODSSGNIDNKPSEENVOSYIKSKDQALNDINEIYVGT 3212
 DB 1575 VLKQLEAKVOEKOKELNENYKKELOELKDVHSHSI-SDDRDKLRA-----EISRLR 1629
 QY 3213 NNENNEPEVKLKYPLGLEYVEDETHKNHVLGNT-----KEEQTVRKALYFTTPHV 3266
 DB 1630 EEFNNELOAIKKRSF-----DEGKQAMKKTLLERLAKMESQLETKQASPEPPS 1682
 QY 3267 DSFYQAPLFSTHVAQD-----PKNDILKSSISVIYVIALGLALHPMKKFKKSVYDLRI 3323
 DB 1683 VNVQNPVLGLPKRIEENSNSPFPNLSGE-----KLKLNKSSS----- 1722
 QY 3324 LNPQEGYCPMTLESKRNY-----PYRSGPYKGYTYVEGDTSCDEKYMWDSSDI 3378
 DB 1723 ---SGGFNFPTSPSPKHKHQNDRKRESLANTPDPTHLPEFPNIPASGL--ISSST 1776
 QY 3379 TSESEYEELDINDIYVPGSPKYKTLLEVLEPSKRDIP-----SDTTPSNDTPPTNR 3431
 DB 1777 LSTDTNDEELTNNPQKOSNNRNOSEEDTEKKKEGEPYKGEALIEQTKSNRNP----- 1832
 QY 3432 FIDDENNELKHDFVSOYLPTPEBNNNKSAIDIMNTEPNTLYSDNPEK 3480
 DB 1833 --IDEVGLKND-----EDDTENIN-----ESKRIKTEDEEK 1864

RESULT 9
 PVDG_PLAKN STANDARD; PRT: 1070 AA.
 AC P50494;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Duff receptor, gamma form precursor (Erythrocyte binding protein).
 OS Plasmodium knowlesi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5850;
 RN SEQUENCE FROM N.A.
 RP MEDLINE=92357776; PubMed=1496004;
 RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
 RT "A family of erythrocyte binding proteins of malaria parasites."
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7089-7089(1992).
 CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH, TO P.YIVAX DUFFY RECEPTOR.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC EMBL: M9095; AAA29604.1;
 CC DR Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
 CC Multi-gene family.
 KW SIGNAL.
 FT CHAIN 1 21
 FT DOMAIN 22 1070
 FT TRANSMEM 22 1003
 FT DOMAIN 1004 1025
 FT DOMAIN 1026 1070
 FT CARBOHYD 134 134
 FT CARBOHYD 134 134
 FT CARBOHYD 179 179
 FT CARBOHYD 676 676
 FT CARBOHYD 743 743
 FT CARBOHYD 785 785
 FT CARBOHYD 936 936
 FT SITE 279 281
 POTENTIAL.
 DUFFY RECEPTOR, GAMMA FORM.
 EXTRACELLULAR (POTENTIAL).
 CYTOPLASMIC (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CELL ATTACHMENT SITE (POTENTIAL).

SO SEQUENCE 1070 AA; 120931 MM; 703D68811BC11B50 CRC64;
 Query Match 1.7%; Score 327; DB 1: Length 1070;
 Best Local Similarity 18.1%; Pred. No. 6,3e-08;
 Matches 240; Conservative 195; Mismatches 467; Indels 424; Gaps 60;
 QY 1014 MKCDIKLKSGHSGHOSQSSCYSDHTPLDDYIIPQKLRMTAEAWCYQKKEVDKL 1073
 DB 35 LECENEYVKNENYKLTATGHH--YMDNDQIE-----RWL-----QGT 69
 QY 1074 ECKEKCCKDNGGCGCTRESGTGCTEACNEYNIDILGKMEQNT--ISDKYKELHEQA 1131
 DB 70 DRSRRAKTEEN-----VKRYVNEELMTKEQMKGK- 100
 QY 1132 QMSVNSNGIEASSTAKNHDNRNIEFLSELYQNGKSNK-----SGTSD 1177
 DB 101 --RINRLKSTYEAQVAVANNYIDKANGKEYKTDRNTKNGEGARNNVMDYDISGSGP 158
 QY 1178 SAYIGTWTYENYAVY-HGTGNFDCQSONERCDEKSDQKDNKYAFRDKPDHDGACG 1236
 DB 159 DGII-----DNVVELLTEDEGNF-----LKNSSKGDHPYMKRKRKSSGA-- 200
 QY 1237 CKSGSKPTVQIKTKKAEKDECTVNDILKENDGKKQVEDCHPKNSNGYPDQCGN 1296
 DB 201 -----INQIFLNNYMDKND-----KRRGERDMDCP 229
 QY 1297 INLVEDPRVCMPPRRQKLCVHFLAN-----DNEIKRLQSVNL-----KEAFI 1339
 DB 230 ---EKDVCIPDRRYQLCWMEITNLVDTPHFSIDIIFRKSYSRRRLIYDVGGRDILL 284
 QY 1340 KSAAEFTFFSWYIYKSKDGEENELDKELGKGIIPAPLBSMFTYFGYRDLFTSDISK 1389
 DB 285 K-----KYNVYSEDCDKDK-----SLQDFGDIIMQTDK-EG 317
 QY 1400 HGECSKLKEQIDSLFKNGDQKSPNGKTRQEWTEHSEIWEAMCALYIGAKKDPTEN 1459
 DB 318 IGSVLVQNNLRISFGTSAELD--RKKWMDHKADIKAMLSKE-----KNR 366
 QY 1460 YGVNNVRESKSTLLEFARPOFLRWLTWDYCYTRQKYLKDVQEKSKNSDOLKCDT 1519
 DB 367 YSAWCK-EDVQIKVE-----PQIYRWIREGRDYMSEFRQRRKLNKEC--DKLYST 418
 QY 1520 -----ECKKCKDYKYKMK-KKEWIPQDKYKDERDKRFRPQIAG--VMYDTY 1567
 DB 419 MLCTLPCCNNACSSYEWETGKKQWDVLTSSVSKKQKLETENIARAYDILKQELN 478
 QY 1568 GTNATDILNKRFTASCGDKPGASVVOIRIOLKQAYYADAKHCGCTFIENDKTYNI 1627
 DB 479 GFNEVTFENE-----INKRDLYN---YF-----CYCIVQEARKNQOE 513
 QY 1628 SSKDCKGLVKEANTGAIKMOKNGPNYNNLKELTEDVLPSPRLRIQF-----AL 1679
 DB 514 N-----YKNVSGV--ESKAPSS-NPINEAVKSSGGEQVEDSAHRSVNEGSKS 561
 QY 1680 DGNTPDEVNDENLRKRMVATVEGYNLQYIKKKEKKEKITSQAKSYEPPCSA 1739
 DB 562 STNADPGSPGPGPASPASVDEKAGVPALISAGQGDKPPAPAAATESAHPASAKTPITA 621
 QY 1740 MKSYFDLRDILIGIDNLEDEKQTEENLKIFKNKGTSVK-----GSDSTTG 1789
 DB 622 -----TEENKQRTQ--VDSVAGDGDKAPRPYVSQVPSVSGKDSGSP 662
 QY 1790 PGSTARFENWNEFCYWNAMICGYKGRDNGSNGSARSDEDLKCGSVPSD--DYPM 1847
 DB 663 PAS-----HLAGNCE-YHNG-----TDTEPKEDGEKADPOKNEIVGKQOTDRSGSL 711
 QY 1848 GKNDEGATVFLWPAWEGDFCKHKELEKLVGAC-----NDYICGDMD--KRRK 1889
 DB 712 GPHYDER-----ASLGE--THMEKDTETGSGTLPEQVNVSAASDGNVPGSGNK 759
 QY 1900 CTDACTQYKPFSEMKPOYEQKIRKYGKGNKDIYSEHPVADADAREYLDKQLKICEN 1959
 DB 760 QNEGATL-----SGAESLSSSESVHTINTHTGL-----EN 792

QY 1779 VKGSGSTGNGPGSTARKFFMNNKCYNNAMICGKRGDNGSNGSARSDEDLKCGS 1838
 DB 655 VG-GKDS---GPTSTASHALAGENG---VANG-----TDEPEDEKDKPPDQIVKGG 703
 QY 1839 VPSDD--DYPGKNGREGTAYOFLRWFPAEWGDFCKHKEKELEKLGAC-----NDYTGG 1891
 DB 704 QPDDSDSGSLGHTHER-----ATLGE---THMEKDTAGSITLTPQNVSVAS 751
 QY 1892 DNEED---KRRKCTDCTOYKKFISEMKPOYKQIKKYGKNDKIYSEHPVAKDAEDAREY 1948
 DB 752 DNGNVGSGNKGNEGATL-----SCAESLKSNESVHKTDIDMTYHG 792
 QY 1949 LKQALKKICENKSGDEYCKMKDVSTORLTDGNSQWMPASLDPEKEVEGCKCQYPRG 2008
 DB 793 L-----ENKNGNE---KDFKHDENNMNDLQASDHTSD-----827
 QY 2009 PVRRETPSPRVS---LISATASKKEAKTAPPTKQPKVENLTJTEBRAQTFR---RA 2061
 DB 828 -----QTSDDHTSSDQTSDDHTSSDQTSDDQTSDDQTSDDQTSDDQTSDDQTSDD 882
 QY 2062 AQQTR--RSTATTES-----DV-GTWKALIS---NKPSRGG 2096
 DB 883 SEDMSKGEFERNNSNELYSHNNLNRRKLNRDQYEHHDYKATREKILIMSEVKKCNRAS 942
 QY 2097 IECNPKTYGQYKMGCIYKSKENE--NGICMPRRKLL-----CINNIQIYLYETE 2147
 DB 943 VKYCN-----TIEDKMLSTCSERRRKNKCCSIDPCNLVFLYLSYEFY 986
 QY 2148 NKDNDIKKFAIKC-----AIEFQFLMKYIIEEN-PAEENELONGTIPPE 2193
 DB 987 NCKKKEFEDSYECFTKGSSTGIYVATGAGFILLILPRASMAASNDYEEATFDEF 1044

 RESULT 11
 MST2.DROHY STANDARD: PRT: 1391 AA.
 AC 008696:
 DT 01-FEB-1995 (rel. 31, Created)
 DT 01-FEB-1995 (rel. 31, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Axoneme-associated protein mst101(2).
 GN MST101(2).
 OS Drosophila hydei (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7224;
 RN [1]
 SEQUENCE FROM N.A., AND CHARACTERIZATION.
 MEDLINE=95045538; PubMed=7957199;
 Neesen J., Padmanabhan S., Buennemann H.;
 "Tandemly arranged repeats of a novel highly charged 16-amino-acid
 motif representing the major component of the sperm-tail-specific
 axoneme-associated protein family Dmst101 form extended
 alpha-helical rods within the extremely elongated spermatozoa of
 Drosophila hydei.";
 RL Eur. J. Biochem. 225:1089-1095(1994).
 CC -1- FUNCTION: POSSIBLE STRUCTURAL ROLE IN THE SPERM TAIL.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: TESTIS. PRIMARY SPERMATOCYTES AND EARLY
 CC SPERMATIDS.
 CC -1- DOMAIN: THE PREDOMINANT STRUCTURE IS ALPHA-HELICAL.
 CC -1- POLYMORPHISM: LENGTH POLYMORPHISMS EXIST BETWEEN DIFFERENT
 CC STRAINS, MOST LIKELY CAUSED BY LENGTH VARIATIONS WITHIN THE TANDEM
 CC REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL: X73481; CAA51876.1; -
 DR PIR: S34154; S34154.
 DR Flybase: Fgno020733; Dhyd\mst101(2).
 KW Sperm; Repeat; Multigene family; Polymorphism.
 FT DOMAIN 332 1268 59 x 16 AA APPROXIMATE TANDEM REPEATS OF
 FT [KR]-K-X-C-X-X-A-K-X-K-X-K-X-X-X-E.
 SQ SEQUENCE 1391 AA; 159000 MW; 1B2A368F30E48878 CRC64;

 Query Match 1.6%; Score 317; DB 1; Length 1391;
 Best local similarity 19.2%; Pred. No. 2.5e-07;
 Matches 334; Conservative 232; Mismatches 660; Indels 510; Gaps 75;

 QY 1507 ECKSNDQKCD---TECNKKC-----EDYVKYMKKKKEMIPDQKYYKDER-----DKRR 1553
 DB 50 ESTDSDNQIICDADCTAADKAKKOKKEBQAKIREYKRECLKQKRVLAEEIRGSEKDR 109
 QY 1554 FDRQHIGVAVTDYTGNNADYLNKFTASGQDP-----GSASVQNRNQLQEK-QAYTD 1607
 DB 110 ILIQQMKCLTDGKKRCKTIAKAKLIA--DKELAVQCAALSKDKVALLKRCERKS 166
 QY 1608 ADKRCGCTFEIENDDKTYTNISKDC-KGLVKEANTGAIKQKNGPNYNNLKELTEDYL 1666
 DB 167 KEKECKNQNSPABD-----KDRTKKGTGKSGSGGKKRSTKENRAKKKLYKNP 218
 QY 1667 FFSRLRLICFHA-----LDGNYTDE-----VKDENGILRRRLM-EVATGEG 1706
 DB 219 --TQKLEHCISQWADVECRQNTFEDERRLAASYCMQTKIKSICRRVIAEMCEAG 276
 QY 1707 YNIGQYKKEKEKEKIKITSNAHKYSYEVPPCSAMKSPFDRLIIGIDNLEDE--KQRT 1764
 DB 277 YKKS---SEKKKKKKKKKDEKKEK-----ELEKEILKEQA 309
 QY 1765 EENLKRIFNKNGTSVGKSDSTGNGPGSTARKFFMNNKCYNNAMICGKRGDNGSNGS 1824
 DB 310 EEEAK-----IRGVKKEVKKKKCKKAL--KKKKKD---LG 339
 QY 1825 NSARSDKDLKCGSVSDQDPYKGNRDEGTAYOFLRWFPAEWGDFCKHKEKELEKLVGA 1884
 DB 340 RKMKKEAEKKKCAALK-----KQKEDEKK---A 366
 QY 1885 CNDYTGDNF-DKRRKCTDCTOYKKFISEMKPOYKQIKKYGKNDKIYSEHPVAKDAE 1943
 DB 367 CKELAKKKKEADKKCKCEEAANKER--AAEKKKCKAAERKEAEKKCKCEAAKKE 424
 QY 1944 DAR-----EYLDKOLKRICENKSGDCYKCMKDVSTORLTDGNSQWMPASLDPEKEVEG 1999
 DB 425 AAEKKCKEELAKNIKKAEEKKCKEAEKKCKEAEKKCKEELAKKIKKAEEK-----K 478
 QY 2000 CNCQVPRGPRVRRTPSPRVSLSKATAS--KKEAKTAPPTKQPKVENLTJTEBRAQT 2056
 DB 479 CEETAKKGEKVAARKECEELAKKIKKAEIKKKCKLAKKEKEFAEKKCKEAKKKEKREA 538
 QY 2057 RTBRADQPTKRTSTATTESDVGTVWKAALSKPPSRGIECNPKTYGQYKMGCIYG 2116
 DB 539 EKKCKEAAKKEEA-----EKKCKEKS 562
 QY 2117 KSKENENGICMPRRKCLCINNIQIYLYETENKRDNDIKFAIFKCAIETOFMLKYIE 2176
 DB 563 AKRKE-----AAEKKKC-----EKAERKEAAKKECEE-----593
 QY 2177 NPAENELONGTIPDEKRIYTYGDYKDMFEGTDSNDKTIYTVNSYTTILNENKK 2236
 DB 594 --AAKE-----KEVAERKCKEELAKKI--KK 616
 QY 2237 KQKKKDEELKRTFWEKNNKFIWEGMIYGLTYLITDENR-----EKIRNYYQNDMTL 2291
 DB 617 AAKKCKCKEAK--KEK-----AAERKCGELAKKIKKAEEKCKCKL 658
 QY 2292 TPLSEFVKRPOFLRWFTEWAEFCNKRKEOLIK---LEAGCKEYECNGSNDKTOCEAG 2348
 DB 659 AKKEKEFAEKK-----CEAAKKRKEAAKKCAEAAKKEKKA-----AEKKCKEE 705

RT "Conserved protein domains in a myosin heavy chain gene from
RT Dictyostellium discoideum.",
RT Proc. Natl. Acad. Sci. U.S.A. 83:9433-9437(1986).
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RC STRAIN-AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lucke-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RA Gerisch G. ;
RT "Replacement of threonine residues by serine and alanine in a
RT phosphorylatable heavy chain fragment of Dictyostellium myosin II.",
RT FEBS Lett. 269:239-243(1990).
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagle G., Noegel A., Scheel J., Gerisch G. ;
RT "Phosphorylation of threonine residues on cloned fragments of the
RT Dictyostellium myosin heavy chain.",
RT FEBS Lett. 227:71-75(1988).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=619793;
RA Fisher A.C., Smith C.A., Thoden J.B., Smith R., Sutoh K., Holden H.M.,
RA Rayment I. ;
RT "X-ray structures of the myosin motor domain of Dictyostellium
RT discoideum complexed with MgADP, BeFx and MgADP.ALFA-";
RT Biochemistry 34:8960-8972(1995).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I. ;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RT truncated head of Dictyostellium discoideum myosin to 2.7-A
RT resolution.",
RT Biochemistry 34:8973-8981(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I. ;
RT "X-ray structure of the magnesium(II)-ADP.vanadate complex of the
RT Dictyostellium discoideum myosin motor domain to 1.9-A resolution.",
RT Biochemistry 35:5404-5417(1996).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gullick A.M., Bauer C.B., Thoden J.B., Rayment I. ;
RT "X-ray structures of the MgADP, MgATPgammaS, and MgAMPNP complexes
RT of the Dictyostellium discoideum myosin motor domain.",
RT Biochemistry 36:11619-11628(1997).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagshaw C.R., Rayment I. ;
RT "X-ray crystal structure and solution fluorescence characterization
RT of Mg.2(3')-O-(N-methylanthraniloyl) nucleotides bound to the
RT Dictyostellium discoideum myosin motor domain.",
RT J. Mol. Biol. 274:394-407(1997).
RN [9]
RP -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
CC -1- SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATOR LIGHT CHAIN SUBUNITS
CC (MLC-2).
CC -1- SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
CC CORTEX.
CC -1- DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LMM) AND 1 HEAVY MEROMYOSIN (HMM). IT CAN BE FURTHER
CC SPLIT INTO 2 GLOBULAR SUBDOMAINS (SI) AND 1 ROD-SHAPED
CC SUBDOMAIN (S2).
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -1- PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND PERTURBS

CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
CC MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2)EDTA ATPASE
CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
CC POSITION (688).
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IO DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL: M14628; AAA33227.1; -
DR PIR: A26655; A26655.
DR PIR: S00250; S00250.
DR PDB: 1MMA; 03-DEC-97.
DR PDB: 1MMD; 17-AUG-96.
DR PDB: 1MMG; 03-DEC-97.
DR PDB: 1MMN; 03-DEC-97.
DR PDB: 1MND; 17-AUG-96.
DR PDB: 1MNE; 17-AUG-96.
DR PDB: 1YOM; 23-DEC-96.
DR PDB: 1LYK; 28-JAN-98.
DR Dictydb: DD01008; mhca.
DR InterPro: IPR000048; IO_region.
DR InterPro: IPR004009; Myosin_N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF02736; Myosin_N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
DR MYOSIN; Coiled coil; Actin-binding; ATP-binding; 3D-structure;
KW Calmodulin-binding; Methylation; Alkylation; Phosphorylation.
FT DOMAIN 1 761 MYOSIN HEAD-LIKE.
FT DOMAIN 762 791 IO.
FT DOMAIN 817 2116 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP.
FT DOMAIN 638 660 ACTIN-BINDING.
FT DOMAIN 738 752 ACTIN-BINDING.
FT MOD_RES 130 130 METHYLATION (DI-) (POTENTIAL).
FT MOD_RES 678 678 ALKYLATION (SH-1).
FT MOD_RES 1823 1823 PHOSPHORYLATION (BY MHCK).
FT MOD_RES 1833 1833 PHOSPHORYLATION (BY MHCK).
FT MOD_RES 2029 2029 PHOSPHORYLATION (BY MHCK).
SQ SEQUENCE 2116 AA; 243871 MW; 2FC3770BB1E556A1 CRC64;
Query Match 1.6%; Score 313.5; DB 1; Length 2116;
Best Local Similarity 18.0%; Pred. No. 6.3e-07;
Matches 377; Conservative 336; Mismatches 827; Indels 551; Gaps 81;

QY 646 IKKLFTRKKNIOOSYSNINNNLEBGFYKMDKLDDEAKWELMIRKKNEFSLN 705
DB 666 -----
QY 706 NRDYLENAEILLDLHLEKATATICKDNNTNEACETSHATNPCKPFGQPPKNIKEA 765
DB 698 YADFEVKRYLLA-----PVPRAEDSQAT--DAVLKHLNIDPEQYRGIT 742
QY 766 QYERKSA---YEARNRGLHKLKGAHEGIYKRGKRDKFKN-----LCRIMIKHSR 816
DB 743 KIFPRAQGLARIEBARQRISEII-KAIQAATGMIAKRYKQAAREHTVAARIIOQLRA 801
QY 817 NLGFSKPCPCGKTGIGCIOTRFVGTW-----EVDEHMKKHEDYIMPPRRRH-CTS 870
DB 802 YIDFKSWP-----WMLFSAFRLKRRNEKEIKEREIELELS 842
QY 871 NLEHLOTHDHPPLNGNIYVDLVNNSFLGDVLLSAKYENKILIRYKKNKNGK----- 924
DB 843 NLDTSTQKRLKSLND---TESNVLDLQRLKAKEETIKAMTDSKDLAQAQRELEIR 899
QY 925 -----EVDPKHQTICRAIRYSFADIGDILINGRDLMERNGDVKLQGHLETFVGNIR 978
DB 900 VEDMESELDKRLALEMIOQKKSVEE-----KYRDLSEELQERQKRLNLEKTKKYE 954
QY 979 SLK--GKNDKYNDAPKYIKLENNMWEANRAKYEMAKCDIKYKKSCHOSTQSYCG 1036
DB 955 ELEBKRVNDGQSDTISRLEKIKDEL---QKEVEELTESFSESKDKGVLEKTRVAL-- 1008
QY 1037 YSDHTPLDVIYIPQKLRMTMAEWYCKVOKKEVDKLEKKECKDKXNGGCGRRSGTGC 1096
DB 1009 ---QSELDL---TVRLDSETKDSKSELRLRK--KLEELKYOE-----ALAAATTA 1053
QY 1097 TKCTEACNEYNDIIGLKEQWNIISDKYKELHEQAOQSV--SNSGIEASSTAKNHDNRVI 1155
DB 1054 KLAQEAANKR-----LOGEYTELENEKFESEVARSNVS---KTLLESQLV 1097
QY 1156 EFLSELYQONGKSKNGSTSDESAIVGTNTTYEVANGVLYHDTGNFDDCOSONECDKSD 1215
DB 1098 AVNNLEDEE---KNNRDLERKKRAL--DAMLEBKQDLSTG--BKSPLYDLKVR 1147
QY 1216 GKDNKAYFRDKPDHOGACGCKSGSKPT---RVQ-----IKTKKAEEK 1257
DB 1148 -QSDMALNRQISELOSTAKLEKIKSTLEGEYARLQGELEAEQLAKSVEKQKKVEL 1206
QY 1258 DTECKTVNDILKENDGKQYEDCHPKKNSNGYPMDCGINIVEDPVCMPRRQKLCVH 1317
DB 1207 DLEDKSAQ--LAETAKQALDKLKK----- 1231
QY 1318 FLANDNEIKKLOQSVNLKEAFISAAAEFTFSWYYSKSGEGNELDKLK-BGKIPPAF 1376
DB 1232 ---LEQELSEVQTO--LSEANNKKNVNSDS---TNKHLFTSFNNKLLELEAEQAKAQL 1281
QY 1377 LRSWFYTFGDYRDLFPTDISKGHSGSKLKEQIDSLFKKNGDKSPNGKTRQEWMTESH 1436
DB 1282 EK-----KRLGLESELKHVNEQLFEKKQKQSEKKRVNDEKE--- 1319
QY 1437 EIWEMALCALVKIGAKKDEFTENGYNNVFRSDKSTLLLEFAKRPOLRLMTEVYDYCY 1496
DB 1320 -----VSELKDQLEE-----EVAASKAVTEAKNNKSESELDEIKRQYADVVS 1360
QY 1497 TRQKYLKDVOEKSSNQQLCDTECNKCKEDYVYKMKKKEMVLPD--KYKDERDKRRF 1554
DB 1361 SROKSVQALTKLQAKNLELRLTAEEAGOLDRAERSKKEAFDELEAVKMLEETAKVR 1420
QY 1555 DROHIGWATVDYGT---NATDYLRRKFTASGCDPGSASVYQVNIQLE--KQAVYDA 1608
DB 1421 AEKAMKAELFDYRSTKSELDAKVSSEY-----VOIRLNEELSELRSVLEEA 1470
QY 1609 DKHCGCTKFIENDKYNTISSKDKGLVKEANTGAIKMONKGPNNYNNLEKTEDEVLPF 1668
DB 1471 DERCNESA-----IKAKTAESALESELDEIDANNAKAKAERKSKELFVRAVELSESLDK 1526
QY 1669 SRRLRICE-----HALDQNTYDPEVK---DENGRLKRLMEVAALBGNVLGQYK 1714

```

Db 1527 SGTVAEVEIRKKAIDIDLRETESRIKSEDEKKNRKPADLEA-----K 1576
QY 1715 EKKKEKIKTSADAKYSEVPPCSAMKSYFDLIDILIDNLEDEQKTEENIKKTFNK 1774
Db 1577 VEEQOREVITIDIRKKLE-----SDIILISTOLDTETTSRIKTEKSKK--- 1622
QY 1775 NGTSVSGKSDSTGPNPGSTARKPFWNENKCVNMAICGKRGHDNSGNSARSDLEK 1834
Db 1623 -----EQTILAE-----RRAAEBSKRAADEIR 1645
QY 1835 KCGSVPSDDVPMCKNDEGTAYQFLWFAWGHEDPCKHKEKEKELVGCANDYTCGNE 1894
Db 1646 K-----QWQOEYDE-----LKAQUDSERAALNASEKIKSLV-----AEVD 1681
QY 1895 DKRRKCTACTOYKFFLSEKMPYOKOIKRYGENKDIYSEHVAKDAEDAREYLDKOLK 1954
Db 1682 EYKQLEDEILAKDLVKA-KRALEVELE---HYRQLEEDSRSLSEKSKRLTTEVE 1737
QY 1955 KIGENKSGDEYKCMQDVSTQRLTDGNSQMPASLDDEPEVEGKCNQVPRPVRRRE 2014
Db 1738 DKKKDAEVEQNTKLDKAKKRLTD-DVDTLKRLQLEDEKKK----- 1777
QY 2015 TSPRVSLSKATASKKAEKATAPPTQPKVENLTTEMRAQOTTRRAAQOTRRK----- 2068
Db 1778 -----INESERAKKRLS-----ENEDFLAKLIDEVANRSRAEDRKRYEKDK 1821
QY 2069 -----TSTAATESDVTGTMVAKILSNKPSNGIEGCPKTYGQYPMKGCIVGSKEN 2121
Db 1822 DPKYKLNDEAAKTKQTEIGA---AKLEDOIDE-----LRSLDEB 1858
QY 2122 ENGICAPPRKRCICINNIQYLNTEENKRDNDKEAFIK-CAAIETQFLMKYIENPAA 2180
Db 1859 QAKATQADSKKTLSEELNLRQIED--EGKIKMLEKREKALBELLELREYEE--A 1914
QY 2181 ENELQNGTIDPEKRIIMYTYGYDKMFPGTDISNDKILITVNSVTTLINENKKNKODK 2240
Db 1915 EDKSEBA---EOKSRILELELDAR-----RMLQKEIDAKLEAEDAKSNQOBEIYAKKR 1966
QY 2241 KXDEELRKLFWEKKNKFTWEGMIGLITLHLDENEKEKIRDNQYNDMTKTLPSSEEFVK 2300
Db 1967 LDEESTAKTNSRSKRRL-EAELIDALTAQV-DAEQKAK---NQOJLENKKIETLEKEYK 2021
QY 2301 RPOFLWFEWMAEFCNKRKEQLL--KLEAGKEYEONGNSNGKQOECACVYONFLK 2358
Db 2022 K-----FGSEKTKKEFLVEKLEETDYKRAKKEADQOOR-----LTVENDLR 2066
QY 2359 KKKTEYEROREKFKDKDQKTKYKDYSPSTERDIEKATCAHEYLNMKIKELCG 2409
Db 2067 KHLSEI-----SLKDAIDKLDORHDKTRKLELETETASKITEMORKMADPFG 2112

```

```

RL Mol. Gen. Genet. 230:277-287(1991).
CC -1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NDM1 SPECIFICALLY CONTROLS
CC THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
CC DIVISIONAL G2 NUCLEUS PERHAPS BY RECOGNIZING G2-SPECIFIC
CC CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
CC ENVELOPE.
CC -1- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT
CC CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH
CC FLANKING DOMAINS OF THE TANDEM REPEATS.
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: X61336; CAA43554.1; -.
DR PIR: S19052; S19052.
DR SGI: S0002557; NDM1.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR KW
FT REPEAT 593 1384 12.5 X TANDEM REPEATS.
FT REPEAT 593 656 1.
FT REPEAT 727 727 2.
FT REPEAT 728 798 3.
FT REPEAT 799 862 4.
FT REPEAT 863 926 5.
FT REPEAT 927 990 6.
FT REPEAT 991 1054 7.
FT REPEAT 1055 1118 8.
FT REPEAT 1119 1182 9.
FT REPEAT 1183 1246 10.
FT REPEAT 1247 1310 11.
FT REPEAT 1311 1374 12.
FT REPEAT 1375 1384 13 (INCOMPLETE).
FT REPEAT 2573 2683 PH.
SQ SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;
Query Match 1.68; Score 309.5; DB 1; Length 2748;
Best Local Similarity 17.98; Pred. No. 1.3e-06;
Matches 600; Conservative 467; Mismatches 1122; Indels 1161; Gaps 164;
QY 95 HTNLRVDVNLRHPCRGREONRDE--DEES-----ECGNKIRYKRNKD 137
Db 3 HNN-RHKKNDDKSSAQYANSIDNSLSQESYSTNGVTRMANIKADECGSGGSDTKRF 61
QY 138 AIACAPRRRHMKDKNLLEALNDINTQNIHDLGNVLTARYEGESIVNHPKGTSDACT 197
Db 62 SISSITSKR-----ETQDVLPFEGAGSSSH-----NGVLTANSKDMFTLESENLLVECR 112
QY 198 ALARSPADIGDIYRGIDMFKNVHDVKE--TGLREYFKKIHDMQMEDEYKNDYDPDSGAY 235
Db 113 KLGSSNNAKNEQIKSLKQIKESISDKIEELTNOKSKSPKELDSTKO----- 158
QY 256 YKLRAMWNVNRRKRWALTCADSYKSGYFMQSESNTPLESNPKCGHKQKQVPTNLDYVP 315
Db 159 -----LN-----MDL-----ESKL-TNLSMRC 174
QY 316 OYLRFWDE-----WGEFCKRKRIKIKAYKDCGRNDEKRYCSHNG-----HDC 359
Db 175 KQLEKAKKTEKSWNDE--KESLKLK-----TDLELITLTKNGMENDLSQKLYHDK 225
QY 360 TTTIKKGLIHLDNK-----CTDCSTCKQYEVWVIGNOEAFFKQKQKYEKEIQSYLS 412
Db 226 EISELKERILDINNENDRLLISVSDLTSELNSIQ---SKRTERIKIQO-----LD 273

```

highly repetitive 313 kDa NDM1 protein.";


```

Db 2032 -ITSEYSELKDID-----RPNDVLKKAAL--FDSIIVE--NIYQOVLNTTS 2078
OY 2329 GCKEYBCNSNDGKTQECACVLYONFIKKMTYERQREKFKKDKGKYDYPSTER 2388
Db 2079 PCBP-----ITVE-----DLKY----- 2091
OY 2389 DIEKATCAIEYLMKKELOGKDCGCKP-----SOLPPTQOOSGDAN 2436
Db 2092 -----AHQF-GMEL-----CLQKPNKLSGAERERIDEQSIMTSSNTTSS 2133
OY 2437 DMPESDIYVEEENKCEC-----PELSKSGMIHTKKTTEPKIP--MNCVEKAAR----- 2484
Db 2134 MTFDALDDNIEELNRYELONNEDYDIISKSSVYKATFTIPAYENIKNSAEKLGKLYP 2193
OY 2485 -----YLSKKAEN----- 2492
Db 2194 FEKSNINLKNIEAPLFKSKNDNDTVASSIDLDHLSRKAKEYGTLLISQDEFEEYHILKDN 2253
OY 2493 --NMOTLKEKPIESTEKESK-----NSWTNNPCDCKRYADPKTYGRN 2539
Db 2254 AVNLNGMEEMNNPLSENONLAKTTNTAOGSAFOPTVPHNDMEVEYGP----- 2306
OY 2540 PCENRENEPKYDYEMKCYKNSKCYOEKRYCVPRPREMCLRNDEIKTERLKDSNYLL 2599
Db 2307 -----DPFTY-----RQLKR-----PAGDRNLILTSREKTLISR--DDNIX- 2341
OY 2600 KAVRTARNE-----GIDILKNSENGCANPI--CDTKY-----SF 2636
Db 2342 -----SOMEAVYGDIDISDFYDESOEIKNDYDIKTQMKGMCLIPESNFVGASYASA 2395
OY 2637 ADLGIVRGTDLRIGGYLPPEIKLYKEEYTKWKRKNKGRKNDYQVFRSAMDA 2696
Db 2396 QDMSDIV-----VLSASY-----YHNLMSPEDEMKKN- 2421
OY 2697 NKNDIMKATCKAPEDAKLFRKGRMDGFE--RITLLIDCKGHHDDPPVDYIQORFRWME 2755
Db 2422 -----CYSNELQAEVKKR--GIQIALITKEDKKG-QATSKHEHYSHKLNKT- 2467
OY 2756 WSEYCKALMELEKFKKSCDCKTSIDCKNDYDENKCEQCTKCOEYKYNFLAKKSLFD 2815
Db 2468 -----STVSTKSGAKGLAEMAATTVAYEDESHPOIEQSHRTNHHKH---KROOS 2516
OY 2816 IOSNKYKELEQPIYTKISTYDHYONFOYKLTFFSECSVEFS----- 2859
Db 2517 LNSNSTSTHTSSRNTPASRRIVASFPMSRAGSASRTSISLOTLASLNEPSIIPALTQYVI 2576
OY 2860 -EYHETSKCLNYKFNENDGSSNIFTYAFEEETPKSKYKACSCGL-----PSKNPLDNCPTD 2914
Db 2577 GEYL-----FKYPRLG-----PFGFESRHERFFVWHPTLLTLVMSASNPILLENAN 2623
OY 2915 QNKDCKEQLQFTFPCSKNDYDNNLNNNAIYVLNSSDCKGVLIPPRRH 2964
Db 2624 TKTGVAITLGEVSAATDPNPYPTGL--YHKSIVTTEPTIKFTCTPDRH 2671

```

```

RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-Keratinocytes;
RX MEDLINE-93346806; PubMed-8345227;
RA Elgart G.W., Stanley J.R.;
RT "Cloning of the 5' mRNA for the 230-kD bullous pemphigoid antigen by
RL rapid amplification of cDNA ends."
RN J. Invest. Dermatol. 101:244-246(1993).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE-Keratinocytes;
RX MEDLINE-91286285; PubMed-1712022;
RA Tanaka T., Parry D.A.D., Klaus-Kovtun V., Steinert P.M.,
RN Stanley J.R.;
RT "Comparison of molecularly cloned bullous pemphigoid antigen to
RL junction plaque proteins."
RN J. Biol. Chem. 266:12555-12559(1991).
RN [4]
RP SEQUENCE OF 1398-1582 FROM N.A.
RX MEDLINE-94280413; PubMed-8010969;
RA Hopkinson S.B., Jones J.C.;
RT "Identification of a second protein product of the gene encoding a
RN human epidermal autoantigen."
RN Biochem. J. 300:851-857(1994).
RN [5]
RP SEQUENCE OF 1650-2257 FROM N.A.
RX MEDLINE-91216368; PubMed-2090522;
RA Overdie K., Kartenbeck J., Stump S., Magin T.M., Krieg T.,
RN Diaz L.A., Franke W.W.;
RT "The hemidesmosomal plaque. I. Characterization of a major
RN constituent protein as a differentiation marker for certain forms of
RL epithelia."
RN J. Cell. Invest. 82:1864-1870(1988).
RN [6]
RP SEQUENCE OF 1777-2704 FROM N.A. (ISOFORM 4).
RC TISSUE-Keratinocytes;
RX MEDLINE-89067122; PubMed-2461961;
RA Stanley J.R., Tanaka T., Mueller S., Klaus-Kovtun V., Roop D.;
RT "Isolation of complementary DNA for bullous pemphigoid antigen by use
RN of patients' autoantibodies."
RN J. Clin. Invest. 82:1864-1870(1988).
RN [7]
RP DOMAINS.
RX MEDLINE-96199235; PubMed-8621649;
RA Tang H.-Y., Chaffotte A.-F., Thacher S.M.;
RT "Structural analysis of the predicted coiled-coil rod domain of the
RN cytoplasmic bullous pemphigoid antigen (BPAG1). Empirical
RL localization of the N-terminal globular domain-rod boundary."
RN J. Biol. Chem. 271:9716-9722(1996).
RN [8]
RP FUNCTION: COMPONENT OF HEMIDESMOSOME PLAQUE. THE PROTEINS
RN PROBABLY SELF-AGGREGATE TO FORM FILAMENTS OR A TWO-DIMENSIONAL
RC MESHWORK. POTENTIAL INTERACTION WITH KERATIN INTERMEDIATE
RN FILAMENTS.
RN [9]
RP SUBUNIT: HOMODIMER.
RN [10]
RP ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1, 2, 3 (SHOWN HERE) AND 4; MAY
RN BE PRODUCED BY ALTERNATIVE SPLICING.
RN [11]
RP DISEASE: INVOLVED IN THE SUBEPIDERMAL BLISTERING DISEASE
RN BULLOUS PEMPHIGOID, AN AUTOIMMUNE DISEASE.
RN [12]
RP SIMILARITY: CONTAINS 10 PLECTIN REPEATS.
RN [13]
RP SIMILARITY: CONTAINS 1 SH3 DOMAIN.
RN [14]
RP SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
RN [15]
RP SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
RN [16]
RP CAUTION: ISOFORMS 2 AND 4 ARE FRAGMENTS AT THE N-TERMINUS.
RN [17]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RN between the Swiss Institute of Bioinformatics and the EMBL outstation -
RC the European Bioinformatics Institute. There are no restrictions on its
RN use by non-profit institutions as long as its content is in no way
RC modified and this statement is not removed. Usage by and for commercial
RN entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

DR EMBL: AL096710: CAB63786.1; -
 DR EMBL: AL096710: CAB63787.1; -
 DR EMBL: L11690: AAA52288.1; -
 DR EMBL: M63618: AAA55606.1; -
 DR EMBL: U04850: AAA57184.1; -
 DR EMBL: X58677: CAA1538.1; -
 DR EMBL: M22942: AAA5538.1; -
 DR Genbank: HGNC:1090: BPAG1.
 DR MIM: 113810; -
 DR InterPro: IPR001101; Plectin_repeat.
 DR InterPro: IPR001452; SH3.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00018; SH3; 1.
 DR Pfam: PF00435; spectrin; 3.
 DR Pfam: PF00681; Plectin; 6.
 DR SMART: SM00250; PLEC; 11.
 DR SMART: SM00150; SPEC; 3.
 DR PROSITE: PS50002; SH3; FALSE_NEG.
 DR Antigen: Coiled coil; Repeat; SH3 domain; Structural protein;
 DR Cytoskeleton; Cell adhesion; Alternative splicing.
 DR DOMAIN 1 1199
 FT DOMAIN 1200 ?
 FT REPEAT 264 2704
 FT REPEAT 366 449
 FT REPEAT 474 567
 FT DOMAIN 581 633
 FT REPEAT 975 1071
 FT DOMAIN 1179 1380
 FT REPEAT 2165 2222
 FT REPEAT 2223 2260
 FT REPEAT 2261 2296
 FT REPEAT 2297 2334
 FT REPEAT 2385 2421
 FT REPEAT 2492 2531
 FT REPEAT 2532 2569
 FT REPEAT 2570 2607
 FT REPEAT 2608 2645
 FT REPEAT 2646 2683
 FT VARSPLIC 1 55
 FT VARSPLIC 363 367
 FT VARSPLIC 457 468
 FT VARSPLIC 688 692
 FT VARSPLIC 867 884
 FT VARSPLIC 1147 1161
 FT VARSPLIC 2427 2440
 FT VARSPLIC 2441 2704
 FT CONFLICT 1998 1998
 FT SEQUENCE 2704 AA: 313084 MW: A7219E687A634A77 CRC64;
 Query Match 1.6%; Score 309; DB 1; Length 2704;
 Best Local Similarity 17.9%; Pred. No. 1.4e-06;
 Matches 559; Conservative 427; Mismatches 1073; Indels 928; Gaps 140;
 DB 814 SRRNLGFSNG-----PCDGGKGTGDIOTRFVVTGTEWEDPERRMRKHEDVIMPPR 863
 DB 207 SSWTSGSSGMSRLTPSVTPAYTPGPPSGLVNPFSSG---VEPNSL---QTLKMQI 258
 DB 864 RRIICNSLEHLOTDPHPLNGNIVDVLVNSFLGDLVLSAKTYANKIIRMYKTEKNLKP 923
 DB 259 RKPLKLSL-----LDONLTEREINKKFFVODLL----- 286
 DB 924 KEVTDPHQTTICRAIRYSFADIGDIIRGRDLERNGDWVKGLOGHLETVGNLHK----- 978
 DB 287 -----KNVDEMOYOLDKTEM--GSDLPVSFVSHLEN-HKNVHRALEEF 325

QY 979 --SLKRGKNDKYNDAPKYLKRE--NWEANRAKYWEAKKCDIKYLKDKSGHSTOSSY 1034
 DB 326 ESSLKEAKISEIGMTAPLKTVAEKLHRLSEYAKLLFSL-----LKNTRNQHHLN 378
 QY 1035 CGYSDHTPLDDYLPQ--KIRMTENAEMWCVYQKKEVDKLAKECKEDKNGGCGCTYE 1091
 DB 379 -----TLHNFVSRAATNELIWNKEE-----EVAAYDMSEBNTIARKKDYHAELMRE 426
 QY 1092 SGYGTCTKEACENYNDIIGLWKEOWNIISDKYKEHLHQAGMSVSGIEASTAANHND 1151
 DB 427 -----LDQKEENTIS--VOELAEOLL-----ENHRA 451
 QY 1152 RNVEEF-----LSELYOQNGKSNKSGTSDESAGVIGTWT 1185
 DB 452 RLRIEVLKLSVWLVAYRAAQTQMSWILQCGQVEQH-----IKENT 496
 QY 1186 TY-----ENGVAYLHDGNGFDGOSQNEFCDEKS-----DGKDNEKFAFRDKPDHDG 1233
 DB 497 AYEPFNDKAEADYLR--NLKDAIQRKYSKCDRSSIHKLEDLVQESMEKEELLQYKS 553
 QY 1234 AGCGKSGKPTRYQITKKKAKEKDECKTVNDI-LKENDGKROVE-----DCHPKKN 1285
 DB 554 TIANLNGKAKTITQLK-----PRNSDCPLKTSIPKAIQDYQIETTYKDECVLANN 607
 QY 1286 SNGYPPDMQ--CGNINLYEDPRVCM--PPRQKLCVHPLAN--DNEIKKL-----QSQ 1331
 DB 608 SH-RAKMKVISPIGNEMAV--PSCFTVPPKNEAV--DLARIDQYQVNTLWHESH 661
 QY 1332 VNLEKAFISAAAEFTFSPWYYSKDGEGNELDK-----ELKEGKIPPAFLRSM 1380
 DB 662 INNK-----SVWSMWTI-----NEIDRIRASVAVSMDIKMLGHEHQVYL 705
 QY 1381 FYTFGSDYRDL-----FGDISKNGHGESKLEKQIDSLFKKNDQKSPNKTKQEW 1431
 DB 706 SNIQSRPELFEDSQSOFVSGSDITQLEKEVAVVCQYQVOLLKSAEREOESVYNTLY 765
 QY 1432 TEHSH-----EIWEAMCALVKIGAKKDFTEYGVYNNVFSQKSTLEEFARPOFLRW 1486
 DB 766 SEVRNRIRLRENEDEDLIRIIRPLERDLHESV----- 799
 QY 1487 LIEWYDYCYTRQKYLKDVQEKCSNDQKCDTECNKCEY-----VYMKRKE 1537
 DB 800 -----FRTEQEKELKLELRK--DDELGIT--NKCEEPFQAASSVPTLSELN 847
 QY 1538 WIPQDYYOERDKKPRPDROHIGVMTDYGTNADYIANKKFLPASGDKGSAVYQNRN 1597
 DB 848 VVQNNMNYSMSTIIDKYVTAISI--YKGYLISFLRK-----TYNLVKNT 895
 QY 1598 QLEKQAYDADHCGCTKEIENDDYTNISKDKGGLVKEANTGAIKWQNGPNNYNN 1657
 DB 896 QAQEAALVKLYETKLC-----EEBAVADAKNNIENIL-----ST 928
 QY 1658 LKELTEDVLPFSRLRICHFALDGNNTDPEVKNKGLRKLMAVYATGYNLGQYKREK 1717
 DB 929 LKQWRESEV--DEKQVHFAL-----EBELQAKAISDEMFTYKXERDLDF--DMWKEKA 978
 QY 1718 EKEKITSDAH--KYSYEVPPCSAMKSYFDLRIIIGIDNLEDEKQTEENLKKINKNG 1776
 DB 979 DQVLRKQNVHVIDNRRLRLEGISLKYRTD--YHPLDMIOQVETQRTI----- 1030
 QY 1777 TSVGKSDSTTGNPGSTARKFEFNNENKECVWAMIGYKRGRRD-----GNSGNSARS 1831
 DB 1031 -----QENQPENSKTLATQL--NOQKMLVSEIEM--KQSKMDEQKVAEQSAIVYKDV 1079
 QY 1832 DLKCC--GSVPSSDDIPMKANDDECTAYOFLMFWFEMGDFCKHKELEKLVGACNDY 1888
 DB 1080 ELQTTTAAVAVDSQKSPYVRRRMOSSADLLIOEFM----- 1115
 QY 1889 TCGDNDEKRRKCT--DACQYKKFTISEMKPYQEKQIKKYGKNDKIYSEHPY--AKDAED 1944
 DB 1116 -----DLRTRYALTATMTQYIKFAGD-----SLKRLKEEVILINSEHPVNIKELEI 1162

QY 1945 AREYLDKQKIKCENKSGDEYKCMKD-VSTQRLTDCGSONMPASLDDEPREVEGKCMQ 2003
 Db 1163 KR-----CKETS---EHGAYSDLRQKATYLENCKITGKISELRMY-----AE 1204
 QY 2004 VPRGPRVREBETSPRSLISKATASKREKATAPPTKOPKVENLTJE-MRACTRR--R 2060
 Db 1205 LKKOKSHVEBELPKVR-----EAAENELR-----KQCNVDEDISQIKRASEKQKR 1252
 QY 2061 AAOOTKRTSTATTETSDVTWYKALISNKPDSRGIGEGCNPTTYGOYPRMGCIYVSKSE 2120
 Db 1253 RELETTIYREKAEARELE---RVROLTIEAEKRAVEE-NLNFNRQJLEENFTRTTLE 1308
 QY 2121 NENGICMPPRKKICINIOY---LNYETENKRDNDIKAFIKCAIEQIFMLKTYIE 2176
 Db 1309 DH-----LKRDLSTLNDLEQKANKLMEELRRKRDNE-BELLKLIKOMEKDLAFQFOVAE 1361
 QY 2177 NPAENELONGTIDPEFKRIMYYTYGDYKDMFGTIDISDKKIITYT---NSVTIILN 2231
 Db 1362 KOLKEKOKTELEARRKITEIYQ-----TCRENALPVCPTQATSCAAYVIGLOQ 1409
 QY 2232 ENKRRKOD-----KKDEELKRIPEKRRKKTITWEGMITYGLTYHLTDENEKEK 2278
 Db 1410 EHKQKAEELKQOYDELTAANKRAEODMRLELYELN-----ALOLEKISSSEKAR 1459
 QY 2279 IRDNYQYNDMTKLTPLSIEFVKRPPFLRWTEWAEFEKCKRRKOLK-----LEAGKE 2332
 Db 1460 L-----LKDRLDETNTMLRCLKLELE-----RKDOAKGYSQOLRELIGROL 1500
 QY 2333 YECNSNDGKTQCAACACTYQNFIKKMTETEYEROREKFKKDKGKRYKDYSTERDIEK 2392
 Db 1501 NQYTGKAEBAEASD-----LKKIRNYQLELESINHEK--CK-----LQREVDR 1544
 QY 2393 ANCAH-----EYLNMKLEKCGNKDCS---CMQKP---SSQLPKYTOQ-----SQSS 2433
 Db 1545 TIRAHVAANKNOHLNSQHSFRDEKELERLQICQKSDHLKEQEKSEHOLLONIKAEK 1604
 QY 2434 DANDPESIDYVPEEFNKC-ECPELSKSGSMITRKITEPKIPMNCVEKAAVYLSKEAN 2492
 Db 1605 ENNDKIQRIN---EELKEKNECAEMLKQVEBELTRONNETKLMQRIQ-----AESEN 1654
 QY 2493 NMDITLKEKFIPIESTKESKSNMTNPNCPKPYAPDKYIGRRNCEENREKRYVD 2552
 Db 1655 ---LYLEKO-----TIQORCALKTQAD---GPKDOLRSTNEH----- 1686
 QY 2553 YEMKCYKNSKFEYOEKRYVCPVPRREHMLRN-----LDEIKTERLKDSYILKMYRR 2604
 Db 1687 ---LHKQTKTEQDF-----QRIKICLEEDLAKSOLVSEFK--QKCDQOITITIQNTKK 1734
 QY 2605 TARNGCIDILKNFNSBNCAMPICDITMKYSPADIGD-IVAGTDMLRIGGYLPRVEIKLY 2663
 Db 1735 EVRNINAEEL--NASKKEKRRGEOKVOLOQAOYQOELNRLKRYODELHLKTEIEQWTHKRM 1792
 QY 2664 KVFETIYKWRNKNKGRNK-----YNYQUTRSAMWANDKRDYKAMTCAPE 2711
 Db 1793 VLFQESKGFQSAEFPKKEKMEKESVITENDISGIRLDFVSLQDEN-----SRQOE 1846
 QY 2712 DAKL-----FRGRMDGEFERTILIDCKGSHKDDPPVD---DYIPORFRW 2752
 Db 1847 NAKCETNIKELERLOQYREMOQOGOMEANHYQKCOKLEDELIAKREVENIKQKMDQ 1906
 QY 2753 MTEMSEYCKALMELEKFKKSCCHCKTSDCKANDYDENKCEOCKTRCOEY----- 2803
 Db 1907 QIKEHEDVLVLLQCEIQK-KSTARDC---TFKPDF-----EMVVKECOHSGELSHNTG 1956
 QY 2804 ---KNFVLMKSLPDIGOSNKKYELQPIYTKISTYDHQVQVQKIMTFKSECVES 2857
 Db 1957 HLHPLPSPLLRWYQ---EPQPLEKMKOHVYQIIP--KEVQ-FQPGAPARLEKESQOC 2009
 QY 2858 FSEVLIHETSKCLANKYENDDSS---NIRYVAFET--PKSYKE-ACSCYU----- 2902
 Db 2010 YSEYFSQTSSELQITPDETNPITRLSEIERDQALNNSBPAPVQACEMELVYKILP 2069
 QY 2903 ---ISKNLJDCNPIDQKDGCKE---LQTFTEGSKNDYDNNJLN 2940

Db 2070 LELAKNKQYDMHTEVTLTKQEKNPVPSA-EEMMLECCRASGGLKNGDPLKGLDEPFFON 2128
 QY 2941 KNAAYVLNNSDDN-----KGYLIPR-----RRHL 2965
 Db 2129 FGDBHACSVRDEDFKFOGLRHTVATROLVEAKLDMRTIEQLRLGLKTVEEVQKTNKFL 2188
 QY 2966 CWRPITAYNYRRKGDG-----LLKKLLTSAFSGQO-----LLGQKYSSE 3006
 Db 2189 TKATSIAGIYLESTIEKESFASAEERITIIDKRVALLAFLEAQAANGFIIDPISGOTYSVED 2248
 QY 3007 EL-----CEFAEK---YSYADSDIIGCTMDMTSLEKIKKIFPENSNEAT 3050
 Db 2249 AVLKGVDPDEFIRILLEKKAAGVSY-----SKRLSYF-----QAME 2287
 QY 3051 NKTWENNRQIWMAMCGKRIATSKYTLDEGWCQLPKDETYQFLMLIEMAKQACE 3110
 Db 2288 NML---DROGKILIDA--QIASGVIDPVRGIRVPEITAIQOGL-----LNNAILQ 2335
 QY 3111 KHVSDSLKTCPCRSNEDNFAA---SELLROPGCONDIRKYSILNLIKNTMELNLIKRY 3167
 Db 2336 FLHESSMTWVFP--NPNNKQALYSSELLRM--CVFDYESQCTLPFGERNISMLNWK-- 2389
 QY 3168 QIKDOSSGNIDNKP--SEENVOSYIKSKDSQCALELNDINEIYVGTKNNEENFEVLEKTL 3226
 Db 2390 --KTHRISVDPKTGSELTYE-----AFQRNLIEKSIYELSGOQYQWKEAM--- 2435
 QY 3227 YPGLYFVEDETHKKNVLDG-----NIKEE-EGTVPRKALYFTTPHYVDSFYQAPLESTH 3278
 Db 2436 ---FFRSHSSHMLTDITGTGLHFNINAEIPEGTIDKALV-----KKYOGILTYLT 2483
 QY 3279 RVA-----QYDPKNDILKSSISVIVSALG--LTLAFHMKKFKSSVDLRLIINIPQGEY 3331
 Db 2484 ELADSLSLRLVPRKD-LHSPVAGYWLTAASGERISVLAASRNLDVRLTALRCLEAQYSTG 2542
 QY 3332 GMPLESKNRRIYPRSGPYKGTIYMEGDTSGDEKYMMDLSSDITSESEYEELDIN 3391
 Db 2543 GI-----IDPLGKTKRYVAELHRGLVDEGF-----AOQLRQCELVIT 2580
 QY 3392 DIYVGSPPKYYTLIEV 3408
 Db 2581 GIGHPIITNKMMSVYEA 2597

RESULT 15
 ATRX_HUMAN STANDARD: PRT: 2492 AA.
 AC P46100; P51068; Q15886; Q9NTS3; Q9H0Z1;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Transcriptional regulator ATRX (X-linked helicase II) (X-linked
 nuclear protein) (XNP) (Znf-HX).
 GN ATRX OR RAD54L OR XH2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), VARIANT S-1860, AND
 RP VARIANT S ATR-X.
 RX MEDLINE-97123494; PubMed-8968741;
 RA Gibbons D.J., Higgs D.R., Bachoo S., Blake D.J., Quarrell O.W.J.,
 RT "ATRX encodes a novel member of the SNF2 family of proteins: mutations
 point to a common mechanism underlying the ATR-X syndrome.",
 RL Hum. Mol. Genet. 5:1899-1907 (1996).
 RP [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
 RX MEDLINE-97386582; PubMed-9244431;
 RA Villard L., Lissi A.-M., Cardoso C., Proud V., Chiaroni P.,
 RA Coliaux L., Schwartz C., Fontes M.;
 RT "Determination of the genomic structure of the XNP/ATRX gene encoding

RT a potential zinc finger helicase.";
 RN Genomics 43:149-155(1997).
 [13]
 RP SEQUENCE OF 860-2492 FROM N.A.
 RX MEDLINE-95179111; PubMed-7874112;
 RA Staylor C.L., Dabovic B., Gulsano M., Gecz J., Broccoli V.,
 RA Giovannazzi S., Bossojaco M., Monaco L., Rastan S., Boncinelli E.,
 RA Bianchi M.E., Consalez G.G.;
 RT "Cloning and characterization of a new human Xq13 gene, encoding a
 RT putative helicase.";
 RL Hum. Mol. Genet. 3:1957-1964(1994).
 [14]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RX MEDLINE-94214473; PubMed-8162050; Villard L., Staylor C.L.,
 RA Gecz J., Pollard H., Consalez G.,
 RA Millaesau P., Khrestchatsky M., Fontes M.;
 RT "Cloning and expression of the murine homologue of a putative human
 RT X-linked nuclear protein gene closely linked to PKAI in Xq13.3.";
 RL Hum. Mol. Genet. 3:39-44(1994).
 [15]
 RP SEQUENCE OF 2401-2492 FROM N.A., AND VARIANTS ATR-X.
 RX MEDLINE-95211835; PubMed-7697714;
 RA Gibbons R.J., Picketts D.J., Villard L., Higgs D.R.;
 RT "Mutations in a putative global transcriptional regulator cause X-
 RT linked mental retardation with alpha-thalassemia (ATR-X syndrome).";
 RL Cell 80:837-845(1995).
 [16]
 RP SEQUENCE OF 1375-2492 FROM N.A.
 RX Pearce A., Chapman J.;
 RL Submitted (DEC-2000) to the EMBL/Genbank/DBJ databases.
 [17]
 RP E2H2 BINDING.
 RX MEDLINE-98167853; PubMed-9499421;
 RA Cardoso C., Timsit S., Villard L., Khrestchatsky M., Fontes M.,
 RA Colleaux L.;
 RT "Specific interaction between the XNP/ATR-X gene product and the SET
 RT domain of the human E2H2 protein.";
 RL Hum. Mol. Genet. 7:679-684(1998).
 [18]
 RP SUBCELLULAR LOCATION, AND ASSOCIATION WITH PERICENTROMERIC
 RP HETEROCHROMATIN.
 RX MEDLINE-20040663; PubMed-10570185;
 RA McDowell T.L., Gibbons R.J., Sutherland H., O'Rourke D.M.,
 RA Bickmore W.A., Pombo A., Turley H., Gatter K., Picketts D.J.,
 RA Buckle V.J., Chapman L., Rhodes D., Higgs D.R.;
 RT "Localization of a putative transcriptional regulator (ATRX) at
 RT pericentromeric heterochromatin and the short arms of acrocentric
 RT chromosomes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 96:13983-13988(1999).
 [19]
 RP DISEASE.
 RX MEDLINE-20213147; PubMed-10751095;
 RA Villard L., Fontes M., Ades L.C., Gecz J.;
 RT "Identification of a mutation in the XNP/ATR-X gene in a family
 RT reported as Smith-Fineman-Myers syndrome.";
 RL Am. J. Med. Genet. 91:83-85(2000).
 [110]
 RP VARIANT ATR-X SER-1713.
 RX MEDLINE-97196774; PubMed-9043863;
 RA Villard L., Lacombe D., Fontes M.;
 RT "A point mutation in the XNP gene, associated with an ATR-X phenotype
 RT without alpha-thalassemia.";
 RL Eur. J. Hum. Genet. 4:316-320(1996).
 [111]
 RP VARIANT JM GLN-2131.
 RX MEDLINE-96224392; PubMed-8630485;
 RA Villard L., Gecz J., Mattel J.-F., Fontes M., Saugier-Verber P.,
 RA Munnich A., Lyonnet S.;
 RT "XNP mutation in a large family with Juberg-Marsidi syndrome.";
 RL Nat. Genet. 12:359-360(1996).
 [112]
 RP VARIANTS ATR-X.
 RX MEDLINE-97467722; PubMed-9326931;

RA Gibbons R.J., Bachoo S., Picketts D.J., Aftimos S., Azenbauer B.,
 RA Bergoffen J., Berry S.A., Dahl N., Fryer A., Keppler K., Kurosawa K.,
 RA Levin M.L., Masuno M., Neri G., Pierpont M.E., Stanley S.F.,
 RA Higgs D.R.;
 RT "Mutations in transcriptional regulator ATRX establish the functional
 RT significance of a PHD-like domain.";
 RL Nat. Genet. 17:146-148(1997).
 [113]
 RP VARIANT ATR-X LEU-246.
 RX MEDLINE-20123062; PubMed-10660327;
 RA Fichera M., Romano C., Castiglia L., Falla P., Ruberto C., Amata S.,
 RA Greco D., Cardoso C., Fontes M., Ragusa A.;
 RT "New mutations in XNP/ATR-X gene: a further contribution to
 RT genotype/phenotype relationship in ATR/X syndrome.";
 RL Hum. Mutat. 12:214-214(1998).
 [114]
 RP VARIANT SHS LYS-1742.
 RX MEDLINE-99347960; PubMed-10417298;
 RA Lissi A.-M., Millan J.M., Villard L., Orellana C., Cardoso C.,
 RA Prieto F., Fontes M., Martinez F.;
 RT "Mutation of the XNP/ATR-X gene in a family with severe mental
 RT retardation, spastic paraplegia and skewed pattern of X inactivation:
 RT demonstration that the mutation is involved in the inactivation
 RT bias.";
 RL Am. J. Hum. Genet. 65:558-562(1999).
 [115]
 RP VARIANT CWS THR-2050.
 RX MEDLINE-99326061; PubMed-10398237;
 RA Abidi F., Schwartz C.E., Carpenter N.J., Villard L., Fontes M.,
 RA Curtis M.,
 RT "Carpenter-Waziri syndrome results from a mutation in XNP.";
 RL Am. J. Med. Genet. 85:249-251(1999).
 [116]
 RP VARIANTS ATR-X E-175; 178-V--K-198 DEL; S-190; P-219; L-246 AND C-249.
 RX MEDLINE-99219535; PubMed-10204841;
 RA Villard L., Bonino M.-C., Abidi F., Ragusa A., Belongue J.,
 RA Lacombe D., Hanauer A., Philipp N., Schwartz C.E., Fontes M.;
 RT "Evaluation of a mutation screening strategy for sporadic cases of
 RT ATR-X syndrome.";
 RL J. Med. Genet. 36:183-186(1999).
 [117]
 RP VARIANTS ATR-X S-179; L-190; I-194; C-246; F-1552; S-1645 AND C-1847.
 RX MEDLINE-20451413; PubMed-10995512;
 RA Wada T., Kubota T., Fukushima Y., Satoh S.;
 RT "Molecular genetic study of Japanese patients with X-linked
 RT alpha-thalassemia/mental retardation syndrome (ATR-X).";
 RL Am. J. Med. Genet. 94:242-248(2000).
 [118]
 RP FUNCTION: COULD BE A GLOBAL TRANSCRIPTIONAL REGULATOR. MODIFIES
 RP GENE EXPRESSION BY AFFECTING CHROMATIN. MAY BE INVOLVED IN
 RP BRAIN DEVELOPMENT AND FACIAL MORPHOGENESIS.
 RP SUBUNIT: PROBABLY BINDS E2H2. BINDS ANNEKIN V IN A CALCIUM AND
 RP PHOSPHATIDYLCHOLINE/PHOSPHATIDYLSERINE-DEPENDENT MANNER (BY
 RP similarity).
 RP SUBCELLULAR LOCATION: NUCLEAR. ASSOCIATED WITH PERICENTROMERIC
 RP HETEROCHROMATIN DURING INTERPHASE AND MITOSIS, PROBABLY BY
 RP INTERACTING WITH HPI.
 RP ALTERNATIVE PRODUCTS: 5 ISOFORMS: 1, 2, 3, 4 (SHOWN HERE) AND 5;
 RP ARE PRODUCED BY ALTERNATIVE SPLICING.
 RP TISSUE SPECIFICITY: URINARY.
 RP DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF X-LINKED ALPHA-
 RP THALASSEMIA/MENTAL RETARDATION SYNDROME (ALSO KNOWN AS ATR-X
 RP SYNDROME). ATR-X IS AN X-LINKED DISORDER COMPRISING SEVERE
 RP PSYCHOMOTOR RETARDATION, FACIAL DYSMORPHISM, UROGENITAL
 RP ABNORMALITIES, AND ALPHA-THALASSEMIA. AN ESSENTIAL PHENOTYPIC
 RP TRAIT ARE HEMOGLOBIN H ERYTHROCYTE INCISIONS.
 RP DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF CARPENTER-WAZIRI
 RP SYNDROME (CWS), AN X-LINKED RECESSIVE CONDITION CHARACTERIZED BY
 RP MODERATE MENTAL RETARDATION, SHORT STATURE, BRACHYDACTYLY WITH
 RP EXCESSIVE SKIN CREASES, AND WIDENING OF THE KNUCKLES.
 RP DISEASE: DEFECTS IN ATRX ARE THE CAUSE OF JUBERG-MARSIDI SYNDROME
 RP (JM). JM IS A RARE X-LINKED RECESSIVE DISEASE CHARACTERIZED BY
 RP SEVERE MENTAL RETARDATION, GROWTH FAILURE, SENSORINEURAL DEAFNESS,

Db 1401 ---RPRTRSAK---AELBENQSRYSKQKKRRIRIKVQ---DSSSENKNSSEEBE 1447
 QY 2138 FFKCAIIEFQFIMLKIYIENPAENELONGTIPDEFKRLMYTYGDKMFGTIDISNDK 2217
 Db 1448 EEEEEEEBEE-----EEEBEEEDENDSKSPGGRKKI-----RKLKND 1487
 QY 2218 KLIJTVNSTTILNENKKKQDKKDELRKIFWEKNNKFIEWGMILGTLHJTDENEKE 2277
 Db 1488 KRTETONML-----KEEERKRRIAE-----ERERE 1515
 QY 2278 KIRNVOYND-----MTKLPSLEEFVRKP-----QFLRMT 2309
 Db 1516 KIREVIEIEDASPTKCPITTKIYVDEDETKRPIYOVHNNMYIKLPHROVDGQF----- 1571
 QY 2310 EAAEFCKRKKQOLKLENGCKEYBCNSDKTOECAE-----ACVY 2353
 Db 1572 -W-DCCESCVKTKKSPSGCITLACMGL--GKTLQVVSFLHTVLLCDKLDEFALVCP 1627
 QY 2354 QNFIRKWTVEYEROREKFKDKDGKKYKDYPTERDIEKATCAHEYLANMKLELGNKDC 2413
 Db 1628 LWTALNMNNEFEKQWEGKLKDE-----KLEVESLAVK-- 1660
 QY 2414 SCMKPSSOLPRTQOSOSDANDPESLDVYPEEFNKCEPELSKSGSMHTKITEPK 2473
 Db 1661 -----RPERSTYMLQRMQ-----DGVM----- 1680
 QY 2474 IPMNCVKAAYLSKEAENNMDITLKEFIPLESTREKESKNSWTNNPCDPKPYAPDK 2533
 Db 1681 -----IGYEMYNLAQGRVVKSRKLEIF-----NKALVDP---GPD- 1714
 QY 2534 YIGRNPENRENRENFKYDEKCKNSKFOEKKRVCVPRR-----EHM 2579
 Db 1715 -----FVVCDESHILKNEASAVSKANSISRRIILGTPLQNNLIEYH 1759
 QY 2580 CLRNDEIKIERLKDSNYLLKWRRTARNEGIDITKNFSENGCAMPICD--TMK--- 2633
 Db 1760 CMVNF-----IKEN--LLSGIKE-FRNRFINPIQN---GQCADSTVWDVRYMKRAH 1805
 QY 2634 YSFADLGDIVRGTDMLRIGGYLP-----VEIKLYVF-EYIYKWRNKNKR 2680
 Db 1806 ILYEMLAGCVQRRDYTALTKFLPRKHEVYLAVRMTSIOCKLYOYLDHLTGNNSEGR 1865
 QY 2681 NK-----YNDVOTFRSAW---WDANRDKDWMKMTCKAPEDAKLFRGRMDGFERITLID 2732
 Db 1866 GAGAKLFDQFQMLSRWHPWCLDLYISK-----ENKGYFDEDSMDEF---IASD 1914
 QY 2733 KCGHDDPPVDYIPORFRMTWEMSEYYCKALMELEKFKSCDHCKTSDRCKND----- 2787
 Db 1915 SDTSMSSSDYTKK-----KKGKKGK--DSSSGSGSDNDVYEVIK 1956
 QY 2788 -----YDEN-----KCEQCKTRCOE-----YKNFVLKWSLFDIQ 2817
 Db 1957 VNNSRSGGEGNVDETGNPNPSVLKEBSKATSSSPAPDMYKDFVTADAEVLEH 2016
 QY 2818 SNKYKELVE---QPIYIKISTY-----DHVONFVOKLTKFKSECVSESEYLHET 2865
 Db 2017 SGRKWLLEFILMAEIGDKVLFQSOLISLDLIEDFILASREKTE-----DK 2065
 QY 2866 SKCLNYKFNENDGSSNIRTYAFETPKSYKACSCTLPSKNPLDNCPTDOKDGCKELQ 2925
 Db 2066 DKPLIYK-GECKWLNIDYR-----LDGSTYASRK--KMAEE 2101
 QY 2926 FTFCSSKNDYD-----NLDNMNAYLVL-----NSSDDNKGYLIPRRRH 2964
 Db 2102 F-----NDETNRGRFLISTRAGSLGILVANRVIIFDASWNPSTYDIOSIFRYVRFQ 2156
 QY 2965 LCTRPITLANY-----RKDGKELIKKLLTSAFSGOLGQKKSSEELCFEAMKYS 3016
 Db 2157 -TKPYVYRFLAOGTMDKIDYDROVTKOSLSFRVYDQOV--ERHFTMNL---TELYT 2209
 QY 3017 YADYSDIINGTMDTSLSEKIKIFETSEATENKRTWMENNROIMHAMLCGYKIATS 3076
 Db 2210 FE-----PLLDDBPNSKKK-----RDIPLMLPKDITLAELOIHKEHIVGHEHDS 2256

QY 3077 KYTLDEGMCQLRDEETNOFLNMLIMAKOACKKHHVSDSLKTKCPRSNEDNFEASELL 3136
 Db 2257 --LLDH-----KEEBELTEERKAAAEYEA-EKKVL--TMRFNIPGTN----- 2296
 QY 3137 RQPGCONDIRKYISLNLIKNTMENLNIKYKOLKDOSSGNIDNKPSEENVQSYIKSKSQ 3196
 Db 2297 LPPVFSNQTPIIPNGLGALSMSN-----QOLED-----LINQREK---VEATNSV 2342
 QY 3197 CALELDINEIVGT--KNEN 3216
 Db 2343 TAVRIQPLEDIISAVWKENMN 2363

Search completed: April 28, 2003, 10:30:09
 Job time : 116.033 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 65.5138 Seconds
(without alignments)
5197.501 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407
Sequence: 1 MGFSCRYFLIKMGNAASSLE.....IDLINDSLVJLNLILYMKY 3542

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
al number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 08
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	4515.5	23.3	2706	2 T28155	variant-specific s
2	4254	21.9	3006	2 T28625	variant-specific s
3	4010	20.7	3026	2 T28431	variant surface pr
4	3489	18.0	3078	2 T28432	variant-specific s
5	3130	16.1	2924	2 T18378	variant-specific s
6	2714.5	14.0	2664	2 T28626	variant-specific s
7	2686.5	13.8	2647	2 T28161	hypothetical prote
8	2677	13.8	2212	2 T28157	erythrocyte membra
9	2303	11.9	2135	2 T14602	variant-specific s
10	1749	9.0	2197	2 B71600	variant-specific s
11	1743	9.0	2182	2 T28634	variant-specific s
12	1689.5	8.7	2228	2 T14029	variant-specific s
13	1574	8.1	1711	2 C71625	variant-specific s
14	1523.5	7.9	2042	2 T18399	variant-specific s
15	1249	6.4	1729	2 T18396	erythrocyte membra
16	811.5	4.2	431	2 F71600	variant-specific s
17	617	3.2	1435	2 A37793	erythrocyte-bindin
18	548.5	2.8	4550	2 T18440	hypothetical prote
19	523.5	2.7	440	2 E71625	variant-specific s
20	520	2.5	3394	2 T18501	hypothetical prote
21	491.5	2.7	3724	2 T18427	hypothetical prote
22	435	2.2	2829	2 A42771	reticulocyte-bindi
23	433	2.2	2523	2 T18477	hypothetical prote
24	423.5	2.2	4981	2 T18489	hypothetical prote
25	421.5	2.2	1979	2 C71622	hypothetical prote
26	420.5	2.2	2401	2 T28676	hypothetical prote
27	415	2.1	6669	2 S55024	nebulin, skeletal
28	405.5	2.1	2269	2 T28677	thorapy protein -
29	404.5	2.1	2510	2 T28160	hypothetical prote

30	403.5	2.1	2657	2 T18497	hypothetical prote
31	384.5	2.0	2573	2 D71614	hypothetical prote
32	381	2.0	3135	2 A48584	transmission block
33	378	1.9	248	2 C71624	variant-specific s
34	368.5	1.9	2485	1 H71621	serine/threonine-s
35	362	1.9	1819	2 A71928	cag island protein
36	358	1.8	1939	2 T18372	repeat organellar
37	355.5	1.8	2295	2 B71621	probable membrane
38	354.5	1.8	1045	2 T18373	erythrocyte bindin
39	353	1.8	3973	2 B71612	hypothetical prote
40	352.5	1.8	1927	2 G64585	cag pathogenicity
41	347.5	1.8	3844	2 T18402	asparagine/asparta
42	347	1.8	4688	2 F82885	hypothetical prote
43	344.5	1.8	1526	2 A45605	mature-parasite-in
44	342.5	1.8	1070	2 T30848	Duffy receptor - p
45	338	1.7	2500	2 G71609	hypothetical prote

ALIGNMENTS

RESULT 1				
T28155				
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragment)				
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)				
C:Species: Plasmodium falciparum				
C>Date: 15-Oct-1999 #sequence_rev15on 15-Oct-1999 #text_change 20-Jun-2000				
C:Accession: T28155				
R:Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.				
Nature 388, 292-295, 1997				
A:Title: Plasmodium falciparum rosetting is mediated by PfEMP1 and requires complement				
A:Reference number: Z20477; PMID:97373957; PMID:9230440				
A:Accession: T28155				
A>Status: preliminary; translated from GB/EMBL/DBJ				
A:Molecule type: DNA				
A:Residues: 1-2706 <ROW>				
A:Cross-references: EMBL:Y13402; PDB:CAA7831.1				
A:Experimental source: strain IT 4/25/5				
C:Genetics:				
A:Introns: 2493/3				
A>Note: R29R+var1				
Query Match	23.3%	Score 4515.5	DB 2:	Length 2706;
Best Local Similarity	31.8%	Pred. No. 5.9e-203;		
Matches 1191;	Conservative 372;	Mismatches 871;	Indels 1315;	Gaps 123;
OY	34	SARVLEERYAKNTRHPK---YAKEHYDSLKGDTRKEF-----RGSPSTPVNK	79	
DB	14	SADVLEKINTGIYNOKEKRYPE--NELKGLISNAIFVDOLRKELNISPGPSD----	67	
OY	80	HNYYPPYPCNLDHKEHTNLRYDDVNLHPCHGREONRFDDESEECG--NKIRYKRNDA	138	
DB	68	-----SCSLDKHFHNTMETEYEGRRKPCYERNEKFSNBEKAGCSGSDIRDGISAG	120	
OY	139	IACPPRRRMKCDKNLEALNDINTQNTHTDLGNVLTAKTGEBSIVNNHPKGT----	194	
DB	121	GACAPFERROMCDRNLETLNKNNTNTHTDLGNVLTAKTGEBSIVNNHPKGSNGKSS	180	
OY	195	ACTALARSFADIGIVGIMFKPNVDKVTETGLREYFKKHQCMEDVEVNDY--NPDSG	253	
DB	181	ICRLARSPADIGDIVGRDMFKPNDAKYEKGLQVYFGKTIYNSLPAPAKHYAHDGSG	240	
OY	254	NYKKLRAMNVRNKNKWEAITCDASYKSYFMQSESNPLF--SNPCGHRKOGKVPNTLD	312	
DB	241	NYKKLRAMNVRNKNKWEAITCDASYKSYFMQSESNPLF--SNPCGHRKOGKVPNTLD	300	
OY	313	YVPOYLWFPDMEGEPFRKRNKIKLKKYKDCGRNDKELKYSCHNCHDCTTTTWKGIHL	372	
DB	301	YVPOFLWFPDMEGEPFRKRNKIKLKKYKDCGRNDKELKYSCHNCHDCTTTTWKGIHL	360	
OY	373	NKCDCTCKCVFEMVWNGNOEAFKROKEREKYSYLSNDNFVNI--NSEYRKQFE	431	
DB	361	LNCRCRGNACSNYTKWIEIQKQPFQKQKRTIMEI-----KIKTNISNENDEKFE	412	

QY 432 KIKETQVATNDTFLNLLNEGYYCKGGLPEKEDITFTNSADKGIPIYSEYQVCPDQGVK 491
 Db 413 NLDKGYASTINTWFLSJLHNGOCODNIDKNNKTNFKNLE---TFPGSYCEACPITIGVK 469
 QY 492 CDGIKYTHSONDRRNVANNEDKPPW-----GVKPTNTVYSGNEGQDITOKLEFNCS 546
 Db 470 CSNEKCTVTEHEW---NSNRLEPTDTSTKLNATNIDMLVNDGIGALIDNELEKNTOK 525
 QY 547 STNYKDNNOKECYKXNDENIRCKLEONTEIN---DNPKIISHNPELWVLYLLRDT 603
 Db 526 YGILKIGIKKQKQCYLN-NIDOKCI--NVMNSGYFDNK--IAFNVLFOHMLRYFVBDH 580
 QY 604 IKWMDKLTCT---NNTTHCHDECNRNCLCPDRWVOKKEEBMSIKKLFTHKKKIQ--- 657
 Db 581 NMLKERRIDVCIKKENINENICIKREKTCCEGVKLEKEMWMDKLNQHYNOKNNHIFL 640
 QY 658 -----QSYSNINLFEFGYFKMDLKDDEAKWELMENIKRKNESNLENNRDYLEN 712
 Db 641 IFTWITGEFKIT--FPNDFKALDDVDIIN----- 669
 QY 713 AIELLDHLKETATYCKDNNTNINACETSHNATNCPVARGSTOPTKINKELOYFKRSA 772
 Db 670 ---VLDITKE-----CODTH----- 681
 QY 773 YBEARHGLHKLKKAHNEGITYKRGGRKDFKDLGRIMIKHSNRMLGFSNGPCKGTGD 832
 Db 682 ---CKI----- 684
 QY 833 GIOTREVTGEWEVDPEHNRKDEHVIIMPRRRHICTSNLEHLOTDHPLNGNIYDOLVN 892
 Db 685 ---EKIRSIDVDLI----- 695
 QY 893 NSFLGYLLSAYEAKKIIRMKERKNLKGPEYNDPKHQTICRAIRSFADIGIING 952
 Db 696 ---KEIIS----- 700
 QY 953 ROLMERNGDKVLOGLHLETFVGNHSLKKGNDKYNDAPKYLKLENNWEANRAKWE 1012
 Db 701 ---WLONKLEV--- 708
 QY 1013 AMKCDIKYLLKDSGHOSSTSGYCYSDHPLDDYIQOKLRMTAEWYCKYQKKEIDYL 1072
 Db 709 ---C---KSHHEDKHEYC-----CDILPKSVDD 732
 QY 1073 KECKECKDKDNGGCKKESGCTCTCTEACNEYNDIIGLMKEQWMLISDKYELHEQAO 1132
 Db 733 EEDDEEV-----DEKE----- 744
 QY 1133 MSVNSGIEASSTAKNHIDRNVLIEFLSELQONGSKSNSGTSDESNAVJGTNTTYENGA 1192
 Db 745 ---ESSQTTKRN-----SOKGTSASCVKACAIIVGV-L 777
 QY 1193 YLHDYGNFDDCOSNECDEKSDGKDNKAYAFKDPQDDHAGAGCKSGSPVQJRTK 1252
 Db 778 QOSNSGSDICNNKNR----- 793
 QY 1253 KAEBKDEKTVNDILKENDGKQVEDCHPKKNSNGYPMQCNINLVE--DPRVCMPPRR 1311
 Db 794 ---KN-----EMGCDKNTEFVDNCGVCMPPRR 818
 QY 1312 OKLVEHLANDNEIKLOSQVNLKEAFIKSAAEFTFSWYYSKDSGEGHDLKEK 1371
 Db 819 KSICIHMLTLEEQTK--NKYQLEAFIKCAAKETNLMKRYKKNKMEABEL--LKKG 872
 QY 1372 IPPAFLSMRTTGDYDFLEGTDISKGHGSGKLKEQDISLEFNQOKSPNGKTRQEW 1431
 Db 873 IPEFMRIMETTFGDFDCLENDMGK--DVDKVKKINKEVENSSKRGFKKIDPEW 929
 QY 1432 TEHSHLEIEMALCALVYKIGAK---KDDFENTGYNNVFXSDK---STTLEEFKAPR 1481
 Db 930 NENGPQJINGMLCALIHADTDSIKND---NYKYEKVTLLARDSNGMTJSEFAKPP 985

QY 1482 QFLRWLEWYDYCYTROKYLLKNDVOKCSND--OLKCDTECNKCEYVYTKMKK--EM 1538
 Db 986 KFLRWLEWYDYCYTROKYLLKNDVOKCSND--OLKCDTECNKCEYVYTKMKK--EM 1045
 QY 1539 IPDQYKDEBRKDRDROHIGVAVDYGTNAITDVLNKKFTASCDKRGSA-----SVY 1593
 Db 1046 NLDKRYKDRKREKNGIDKGIILIVADYLANKEKLEKFKFTASCVTSSGKONSATEEV 1105
 QY 1594 ORNOLLEKOAYDADHCCGCTFIENDKYTNISSKDKCKGLVKBANTGAIKWONKGN 1653
 Db 1106 KKNLELSEGOYIDAQYCCCTFI--HDKYSKISGNSCCGLNSDPAKKNKIKWRNSDK 1164
 QY 1654 NYNNLKE--LTDVLEPSRLRJCIFHALDGNITDPEYKDENGKRLKLEMYAATEGNLQ 1711
 Db 1165 DYAFKRRNLSGVDFPSPRRLRJCIFHALDGNITDPEYKDENGKRLKLEMYAATEGNLQ 1224
 QY 1712 YKREKKEKETEAKISTDAKYEYVPCGAMKYSFYDLIDILGIDNLEDEKOKTEEMLKRI 1771
 Db 1225 YKREKKEKETEAKISTEYVQPCGAMKYSFYDLIDILGIDNLEDEKOKTEEMLKRI 1284
 QY 1772 FNNKGTSGKGSSTGTPGSTARKEFPFNMENKCECVNAMIQYKRGHDGNSGNSARSDE 1831
 Db 1285 F-----KSENEGSGQROTPEMNNKCGVWEAMKCGYKHGRDGNNSANSASDD 1332
 QY 1832 DLKCGSYPSDDYTPAKKNDEGTAQOFLRPFAMGSDPCKHKEKLEKLVGACNDYTCG 1891
 Db 1333 DLKCGSYPSDDYTPAKKNDEGTAQOFLRPFAMGSDPCKHKEKLEKLVGACNDYTCG 1392
 QY 1892 DNEKRRKCTACTQYKRFISEMPPQYKQIKYKGENKDKIYSEHPPAKDAEDAREYLDK 1951
 Db 1393 DNEKRRKCTACTQYKRFISEMPPQYKQIKYKGENKDKIYSEHPPAKDAEDAREYLDK 1452
 QY 1952 QLKRICENKSGDCEYCKCKQDVSTORLTDGNSQNNPASLIDPEKVEBKCNQVPRGPPV 2011
 Db 1453 QLOKSC--NSGCKD--CWNKKT--SNGCNPASLIDPEKVEBKCNQVPRGPPV-- 1502
 QY 2012 RRETPSPVSLISKALAKKEAKT-----APPTKOPK----- 2044
 Db 1503 ---PPAPRPPAPRSGNDYGRSEPGDGLPLPPLPPPPKGGGAGRILSTPRNG 1559
 QY 2045 ---VENI/TEMBAQOTRRAAOOTRKRSTATTESDVTWKAILSKPDSRGIT--BGN 2101
 Db 1560 TIEDEEDDDEBEKAAEASETESK--EVEKEDTEFEKCKIVANILTKGMDACN 1617
 QY 2102 PRTYQOYK---WGCIV-----GSKENENGJCMPPRRKLC 2135
 Db 1618 OK-YG-YPOHMGWKCVPPTSSSTSERGASRNKRNIDSTKSSDKNGSICIPRRKLY 1675
 QY 2136 INNIO--YINYTE-----NKRDNK-----EAFKCAIETQFLMT 2171
 Db 1676 IKKIDEMASGNTQAGNGTSGDSTGASSNPQONGVSTSPVALLHAFAVESAAVEFLMD 1735
 QY 2172 KY-----ITENPAE-----NELONGITPDEFRIMYYTYGDKYDMFEG 2210
 Db 1736 RYKKEKEIEKQOQOTGLVASETSEDTEHPONKLOSQITPDLFQMKYTLGDRDIOVG 1795
 QY 2211 TDISNDKILTYTNSVTTILNENKKNOD--KKRD-----ELARKIPEWKNKKFTMEGMI 2263
 Db 1796 ---KTPDGDITVSASDKDTMDKIQAKIQILPKKTPPSVKTPOQTMWNNKAHESIMNMI 1852
 QY 2264 YGLTYHLTDENKIR-----DNTQYNDMTKL----- 2291
 Db 1853 YALTYK--TTPSGEKKQJPEVTKLDEKGTPOGNNKYQYKVY-KLEEBETSGAPKSTE 1910
 QY 2292 ---TPS-----LEEFYAKRPOFLWFTBMAEFCNKKKBOILKLEIAC-----KEYCNG-- 2337
 Db 1911 SSSPSGENTPLNDIFSRPPEYFYLEWGNFCKEKRRKRLGDISKCLDGDQYQSDGE 1970
 QY 2338 -SN-----DGTOCAEACVYQNFIKKWTREYEROREKPKKDKGKKYK---DY 2383
 Db 1971 CSNIDVNDKIFADLEGPCKAPCSSYRKWITGKTYQYKQKRAEQKQCKKTTNNHY 2030
 QY 2384 PSTERDIERATCAHEYLNNKLKELGKNKDCSCMQKPSQLPPTTQOOSSDANDMPESLD 2443

Db 2278 -----REDWKC-----GEYGVDCICTLEKTNNEERY 2304
QY 2735 GHKDDPPVDYIPQRF--RWMTENSEYYKALMELEKFKKSCDCKNDYDEN 2791
Db 2305 SAKENDNKNQIILIRVLEFKRL-----ESLEEDYKNTKNDKISHC-----MKND----- 2347
QY 2792 KCEOCKRQOEYKFNVLKMKSLFDIOSNKKYKELYEQPIYKISTYDHVONFVKLTFS 2851
Db 2348 KKSQCINGCQNCCKYKNTKKEKSKMGVREHY-----INQY----- 2385
QY 2852 EGVSEFSEYLETSCMLTKFNENDGSSNIRYAEETPKSKYKSCSLPESKNPLDNC 2911
Db 2386 -----RDKNNSN--EAFE--VKSFLETLIPQIP-----Y 2409
QY 2912 PTDONK--DGCKELQTFPCSKNDYDNNLDWMAYLVVLSNDDKNGVLIIPRRHLCSTRP 2970
Db 2410 YVDKGHDSITQLKLLKCKCKSEKSEN-----SNEKDYV-----LC----- 2445
QY 2971 TAYNRKKGDEILKKLLTSFASQGLGQKYKSEELCEAMKYSYADYSDIIGKTDMM 3030
Db 2446 -----LLK----- 2449
QY 3031 DTSLSEKIKKIFETSEATENKRTWMENNRQIWMHMLGCKYKLTATSKYTLDEGWCQLPK 3090
Db 2450 -----LEKAKN-----C--KD 2459
QY 3091 EETNOFLRWLEWAKOACKKKHVSDSLKTCPRSNEDNEFASSELLRQPCQNDIRKYS 3150
Db 2460 QASGE-----PCQOTSSENDDDLLE----- 2483
QY 3151 LNLTKNMENTNIRYKQKLDKQSSGNIDNKPSEENVOSYKSKDSOCALINDINEIVTG 3210
Db 2484 -----ENPVEAPNI-----CPKYEPEPVV--EEEKC-----DLAEPK 2516
QY 3211 TKNNENNEFEVAKLYLPGLYEDEDHKNHVLGDKNKEEQYVRKALYFTPHDSFY 3270
Db 2517 ESSTEENSGE-----GSNSEONPSKPEEPPPSSETDTP----- 2553
QY 3271 QASLFSTHRAQYD-PKN--DILKSSIVIVSALGLALHFMKKFKSSVLDRLNT 3326
Db 2554 -PPAPPLIQSQADQPTNSLSDLSSTIPRGIALALSYFLFKKTKSVDLRLNT 2612
QY 3327 PCEYGMPTLESKNRYIPYRSQPKGYTYMEGDTSGDEDKYMDLSSDITSSSEYE 3386
Db 2613 PKGEYMPPLKSSNRYIPYASDRYKGYTYMEGDS--DSGHYED--TDTVSSSEYE 2668
QY 3387 ELDINDIYVPGSPKRYTLTEVLEPESKRDLPD--DTPS-----NDPTIRKFLIDEM 3437
Db 2669 ELDINDIYVPGSPKRYTLTEVLEPESGNNNTASGKNTPSDPTQNDIQNDIGPSSKITDNEM 2728
QY 3438 NELKHDFVSQYLPNTEPN--NNYKSAIDIPMTEPNPLYSDNPEEKPFITISDRDLYTG 3494
Db 2729 NTLKDEFISQYLOSQPNVNDYTSNGSNSTNTNITTSNHNVEKPFIMSHIDRLNLYTG 2788
QY 3495 KEISYININNSTNTNDIPMANARDSTRGIDLINDSL 3550
Db 2789 EEINYNVNM-VNTMDIPINRDNVNSGIDLINDAL 2823

RESULT 3
T28431
variant surface protein 1 homolog A4VAR - malaria parasite (Plasmodium falciparum) (frag
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28431
R:Smith, J.D.; Kyes, S.; Craley, A.G.; Fagan, T.; Hudson-Taylor, D.; Miller, L.H.; Baruch
Mol. Biochem. Parasitol. 97, 133-148, 1998
A:Title: Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte m
A:Reference number: Z20486; MUID:99094502; PMID:9879893
A:Accession: T28431
A:Status: preliminary; translated from GB/EMBL/DBD
A:molecule type: DNA

A:Residues: 1-3026 <SMT>
A:Cross-references: EMBL:L42244; NID:g3540144; PID:g3540145; PIDN:AAD0351.1
C:Genetics:
A:Gene: var
A:Introns: 2906/3
Query Match 20.7%; Score 4010; DB 2; Length 3026;
Best local similarity 29.8%; Pred. No. 3,1e-179;
Matches 1095; Conservative 480; Mismatches 1151; Indels 954; Gaps 127;
QY 30 ESHKSAKNVLEKAKNI-RHPSKYAKEHVDLSKGLDTLAFEGRGSPVPVKNHNYTPYC 88
Db 13 EDDKDAKHVLDISGEYKKEKRYAANYSSQLGTLSNAIFENEPKQOQTEND-----PC 67
QY 89 NLDKHEHTNLRVDVNLRHPCGCHREONRPEDESESGN-KINRYKKNALACAPRRR 147
Db 68 KLIYEHTNVTQGH-GREHCRKRGTERFSDVGGECDDNKIKD--SKNNGCACPRLR 124
QY 148 HMCNKLALNDINTONIHDLGNVLTAKYEGESTIVNHPH-----GTSDACTAL 199
Db 125 HLCVRNLNISALDKINNDTLADVCLALHEGQSTQDYQYQAYASSFSQICTML 184
QY 200 ARSFADIGDVRGIDMKRPVNDK--VEGLRBYFKIHGMD-EKKNYNDGSGNRY 256
Db 185 ARSFADIGDILRGKLDLYGKKEKLEKLLKIFGKLYEKLDPRAKHDKDEPNNEF 244
QY 257 KLRKAAVNVNRKRYWEALITCDASYSYGFMSQSNTPLPFNPCKGHCKGK-----VPT 309
Db 245 QLRKEDMNNARQVWYKALITOHAGESDKYFRKTAAGTCTGOGRCNDKPKPSNDPPT 304
QY 310 NLDYVQYLRWDEWGESECKRNRIKLYKVDSCR--NDKERLYCSHGHDCYTTWK 365
Db 305 YFDYVQYLRWDEWGESECKRNRIKLYKVDSCR--NDKERLYCSHGHDCYTTWK 363
QY 366 KGLIHLNDKCTDSTCKVYEWLGNQGEAFKQKKEVEEISYLSNDKPY-----NNI 421
Db 364 KHYLMDNACAGCFSSQSDRKMIAOKDEFEKKNKYDEIK--KNDDTTTEYGTI 420
QY 422 NSEYKQFYELKEQYATNDTFLNLTNECKYKCG--GLPBEKDTLPNSADGKIFY 477
Db 421 NNNYRDFYKHLDE-KYKTDALNLTNKEKECKNHEVEBGKKYIDFNDNE--TFS 476
QY 478 RSEYQVCPDQGVKCDGIKYTHKSDNDRERVN--EDYKPPWGVKPTNITVLSG 530
Db 477 HTEYREPCWPGIE-----EOKDKWKRIINDHSAKBEELIYTPKNAKYKTNVLTSG 529
QY 531 NBOGDITQKLENCNSSTNKDKNN-----QKWCYKDE----- 565
Db 530 EGHEDIARKLKECTCTQONGGSSDGCNGSSDLSLCEPMQCYOPDLEKVGGEVDDK 589
QY 566 -----NINCKLEQNTENINDPKIISFNPFEIATVYTLRTIKNDKLTCTCIN--NTT 618
Db 590 GAGGLCIEFKMKKEKKYK-----KQKTENNPFNFWAHYAKOSIDMRQTLTCSIDLK 643
QY 619 THCIDECNNCLCPDRWVYKQKEEWSIKKLTFTKKN-IOOSYYSNNINLPEGYFERVMD 677
Db 644 KKCEKGSKSCQCFKWKIEKKKEKEMIKVQKQTFLEMKHYLVLETILENYEENIQ 703
QY 678 KLDKDEAKKLEMBNIRKKNKNEFSNLENNRDYLENALELLDLHKEATIC-----KDN 752
Db 704 KAYGDLKSIOEMKMKIENQKNKRTKD--EDALDVLDEHKEEAEDCLIHEDDD 759
QY 733 TNBACETSHNATNPPCVKPRGTOPTKNIKEIAOYEFKSAVEBARNR--GLHLKKAHE 790
Db 760 DDCVEIEIEKIPNNPC--SGTRHRAMVAVNAADMYRAARQQLRNAGAKRTLRADASQ 815
QY 791 GIYKRGGRKDFDNLCRIMIKHSNRLNLFSGNCPDCKGTGDTGTRFVVGTEWEDPEH 850
Db 816 GHYNGKANESVULD-VCDIINOYSNA-IGDSKPCNCKGSG-----FKIGTPTWNTIVK 867
QY 851 MRDHEVIMPRRRHICSTNLEHLOTDRHPLAGNIVYDILVNSFLDYLKSAFYANKI 910
Db 868 KTSYKQVFLPRRREHWCSTNLENDVGDVTNNVN-----VNNKFLVALLSANKQAEWI 922

QY 911 IRMTKE---KNNLGPRKVEYTDPKHOTTICRAIRYSPADIGDILINGRDLWERNGDVMTKLG 967
Db 923 KOKYNEBNGONNHKG-----CRALKSSFADLDGILKGTDLMDKDSQEQTOR 970
QY 968 HLEVPFNGHSLKGNKNDKYNDAPYLIKIREMWEANRAKWEAMKODIKYIKKSGH 1027
Db 971 NUVTFPKIKYORKGIDTSKYTNLNDGKHNOJREDWMEANROVWAKKALK----- 1022
QY 1028 QSTOSSYCGYSDHTPLDYLIPOKLRMTWEMAWEYCKYQKREYKIKKECKECKDKONG 1087
Db 1023 --GERKINGC--APPYDDYIPORLRMTWEMAWEYCKYQKREYKIKKECKECKDKONG 1072
QY 1088 CTKESSGCKRCEACHEYNDIIGLMEQNNIISDKY--KELHEQOMSVNSGIEASST 1145
Db 1073 ICNKVNDCKACKTEACEYKTKIOPMKDQWEKLELEYALSYLH----- 1115
QY 1146 AKNHIDR-----NIEFELSELVQO--NGSKSNKSTSDSESAVIGNTTYENVGAY 1193
Db 1116 AKNDSRMARGTDPDYQOVVHFPEKLOEAIKSTSRKRPRSTDAITTPPTPYSTAAGY 1175
QY 1194 LHPTGNFDQOSQNEFCDEKSDKNEKYAFRDKPODHGACGCKSGSKPTVQIKTKKK 1253
Db 1176 IHOEIGNAGCOIQKHCDDBKD-----KYVFERKPRDHDEACCTENVE-----KPK 1222
QY 1254 ABEKOTE--CKTVNDILKENOGKQVEOCHPRKKNNGYPD--WOC--GNINLVEDP 1303
Db 1223 KEEDIDKLDVCAIVKALTATTOJMLTKA--CQOK--YGHPRHMGKCISETTKSSDSG 1277
QY 1304 RVCMPPRROKLCVHFLAN-----DNEIKKLOSOVNLTKA 1337
Db 1278 SLICVPRRRKRLVYTPLTWAEATEBEPSPQAGEATLTPPATASQAKGSL--LTLA 1335
QY 1338 EIKSAAAEFFPSYTYKSDGSENEIDKLEKSGKIPPAFLRSMFTYFGDYRDLFTGT-- 1394
Db 1336 FIOSAAVEFFELMHKKY--MDNNGDAEDKIKGEIPEEFKROMFTLADYDIOICGVKMD 1394
QY 1395 ----DISGHGEGSKLKE--OIBLFFNGQKSPNG-----KITQEWMTESHEI 1438
Db 1395 VIKALEASSDNKSGNIKIKISDKIEIILKQSGSKLPGGLVTPNNVKNKRTWMDNAKHI 1454
QY 1439 WEAMLCALV---KIGAK--KDDFTEN-----YGYNNVFS----- 1468
Db 1455 WGMGMCALYKEDTGGAKKTSITODPRAYGKLMNDGKPKREDYDKIYIVISSVPSN 1514
QY 1469 --DKSTLEEFARPOPLWLEWYDDCYTQOKLXLOVECK-----SNDOL 1515
Db 1515 VPSGDTKLEEFRRPTFFRMLEWGEFEORRKKDKLEIKERCYGLNRRHNYCSGDGH 1574
QY 1516 KCD-----TECNKCEYVYKYM--KKKEMIPQDKYKDERDKRFRDROI 1559
Db 1575 ICEKTDTSNNNTFIDLHCPRLKECITIKRWIEKKEKEFNHOKNNYEKEFN--DLK 1631
QY 1560 GVMWTDYGTANTDYLNRKFTASGCD--KPGASVYQNR--IOLLEKQAVYADKHC--GCTR 1616
Db 1632 G-----YSEFN-----NPLASLNHCHGHEHNRDKNKIEFNHNTKTFGSEYCKAPV 1679
QY 1617 FIENDKTYNISCKDKCKGLVEANTGAIKWONGPNYNYMLKELTEVYLPSSRLRCF 1676
Db 1680 YGVKCNK-----KNGECTIKHTDLNG-----QNDNNYTDIKAYLVD----- 1716
QY 1677 HALDGNATDPEVKBNGILRKLMEVAATEGYNLQYKKEKKE-----KIKTSPA 1729
Db 1717 --RKGESNDELKVCNNTSLFKDSSV-----QYMKCQKKNNEVDCIINFIDIDIK 1767
QY 1730 YSYEVPKASAKYSFYDLROIILIDINLEBKQTERKLIKFNKNGTSVGKSGSDSTGN 1789
Db 1768 Y-----MEFVFPQRMRLRYFVHDYNIILDKIKPICIK-----TKDEKSNKJNGC 1811
QY 1790 PGSTARFEFNEKCEYWNAMICGYKRGDRDNGS-----NSARSEDJLKKCGSV 1839
Db 1812 KGRLECYKMWLQKQDKEMKNIKDIEKKNKSLYGIGFHWKYSYFVEQLYFDKDKYKADY 1871
QY 1840 PSDDDYPMKKNDEGTAYOFLRWFMEWGEDFCNHEKELEKLVGACNDYTCGDNEDKRRK 1899
Db 1872 IED-----ENERKIKMGCTDGEVCTNEEFK 1899
QY 1900 CTDACTQYKKEFISEMPOYKQIKKYGENKDKIYSHHPAKAEDARELDQOLKICEN 1959
Db 1900 --DEFINILKIKLOEKIESQOTQHPNG--KTPPCDIPHSO-----ETLDSOTPTTDD 1950
QY 1960 KSGDCEYK-----CMKVSTORLTDGNSONMPASLDDEFEVEGKONCQVPRGPRVRE 2014
Db 1951 DMSDKIYDTPPCPRDYVDTKET-----EKRP-----VLGPP----- 1984
QY 2015 TSPSPVLSKATASKEAKTAPTKQPKVKNLUTEMAAQTRRAAOQTRKRTSTAYT 2074
Db 1985 ----- 1984
QY 2075 TSDVGTWVAILSNKPSRGGIEGCPNTY--GOYPRMGCIYKSKENEGICMPRR 2131
Db 1985 --DACEIYGEIL--NGODQTKIEBCNTRKYPTKNDYQWNC--TDKVINREBSCMPRR 2039
QY 2132 KRLCINNOIYLYEYENKRNNDIKFAFIKCAIETFOFLMKY-----IEN 2177
Db 2040 QKLCIHNLEH--SEKATELRAKAFTECAAIETFWMLDMKYKEDKDEKKEGGISDD 2096
QY 2178 P--AAENELONGTIPDEFKRIWYTYGDKDMFEGTDSNDK--KITVTNASTTILNENK 2235
Db 2097 PDDPKLEGGITPDEFKROFYTGYDRLFTGDISKHKESALGKIDSLFKNGDQ 2156
QY 2236 KKODKKDEBELKTIWEKKNKFIWEGMITYGLTYHLDENEKIKIRDNY--QYNDMTKLPS 2294
Db 2157 KSPSGKTPLE--WMNDYGPDIWKGWVCGLSHIKIGN--KEQJRNKLTDNKTKTSSK 2211
QY 2295 LEEFYKROPLRMTWEMAWEYCKYQKREYKIKKECKECKDKONG 2354
Db 2212 LEDPFSRQOFLRMTWEMDQOCREYVAKINOLKGCMEYEGSEONKKECKKACACEAYK 2271
QY 2355 NFKRWKTEYRORKEFKKDKGKKYKDYPTSTEDIEKATCAHEYLMMKLIKELGNNDCS 2414
Db 2272 SWLKQMKOYBOQAKAFDKDKKFF--DGISAEDVANAVSVEHYLDEBLKNTLTGDC 2330
QY 2415 CMOKPSSOLPRTQOOSOSDANDPESLDIYPEEF--NKCEPELSKSGSMITHKITEPK 2473
Db 2331 CMERPSAO--DETELLGNGYPEAMDYPPKIGERCKO-----AIPS 2371
QY 2474 IPANVEKAAVYLSKEAENNDITLKEKFIPIESTKESKSNSTNNPCDPKPYAPDK 2533
Db 2372 EPMSCVBOJIAHLKEKAKVKIYESS--IKOTPAK--SKNDCTKID--EAIKGDNGSK 2424
QY 2534 YIGR-----RNCENRE--NRFVYDEMKCYKNSFYQEKRYCVPRRRHMLRN 2583
Db 2425 IINSLIDSTFASCEQSEKDATDRLKIGKOWF--NKINGETELKLYPPRRKDCOFND 2481
QY 2584 LDEIKIRLDSNYLQWVTRTARNEGIDIIKPNSENGBAMPIDDTMYSTADIGDY 2643
Db 2482 LKNOFNEVODSNSLEKIDHAKNBSGIDLKLNLPDQNAFESIDAMKYSPADIGDII 2541
QY 2644 RGTMDLRIGGLPPEVETIKLYVEEYIYKWRNKNKGNKKNYDQTPRSAMWDRNDIK 2703
Db 2542 RGSKID--PTNNKIKELOKIFKOJODNALSUK--ELPELRKEMWADRNEVWN 2595
QY 2704 AMTCKAPEDAKLPRKGMDFEITLID-----KCGHDDPVDYDIPORFPMWTESE 2758
Db 2596 AMTCVAPNDHLKRNKNNPNKSOIIASQTEQTKSHDSEPPDYDIPEBRYFLDGESE 2655
QY 2759 YKCALMELEKFKKSCDHC--KTSDBCKNDYBCKEOKCTROKKNVLYWKSIFDIQ 2817
Db 2656 YKCALKEKNDKNCSCISGATCEKEDEKKECKEENDCKEYKKNVLDWQOSEFDQO 2715
QY 2818 SNKYKELYEOPITYKISTYDHYON--FYOKLTKFKSECS--VESFSEIYHETSKLNYKRN 2874
Db 2716 NOLYKRLYTQDRTGHPSTARRNPSIEFTOKL--EDSCNDPYSADKYLDISHCIDYKFS 2772
QY 2875 ENDGSSNIRTYAFEEIYPRKYKACGCTLPKSNPLDNCPTDNKCKELQTFPFGSKNDY 2934

Db 2773 EFD--SRESNVAFSPYKRYKENCCKV-----NTFF-----SNND- 2806
 QY 2935 DNNLDMNVAIVLNVSSDDKGVILPPRRRHLCRTPTAVNRYKDKKLLKLLSAFSQ 2994
 Db 2807 -----PK-----S 2809
 QY 2995 GOLLQOKYSEELCEFAEKYADYSDLIKGTDMMDTSLSEKIKKIFETSMATEENRT 3054
 Db 2810 PSLIGPSF----- 2817
 QY 3055 WNNNRROIWHAALCGYKATSKVTLDEGCOLPKDEETNOFLRMLEWAKOACEKKHV 3114
 Db 2818 -----FLPK----- 2821
 QY 3115 SDSLTKCPNSMEDNEASELLRQPCQNDIRKYISLNLIKNTMENLIKRYKOLKDOSS 3174
 Db 2822 ---KPK-----MKFYPRIGIGVJHPFIM----- 2842
 QY 3175 GNIDKRPSEENVQSYIKSDSCALELNDINEIVGTCKNNENREKREVLKILYPGIYFE 3234
 Db 2843 -----VADPTIHETVAKTFNNVPOF----- 2864
 QY 3235 DEHKHNVLDGNKEEPTVRKALYFTPHVDSFYQAVLFSTHRVAYOYPRKNDILKSI 3294
 Db 2865 -----HINPKDIDVAP-----TKNT-----LNEVLPSAI 2889
 QY 3295 SVVIVSALGLALHPRKKKSSVDLRLILNPOGEYGMPTLESKNRYIPYSGPYKGR- 3353
 Db 2890 PGIALLAGSIAFLFKKTKHPVDFSVINIPKSDYDPTLSPRRYIPYSGKRGNG 2949
 QY 3354 TTIYVE--GDTSGDEKYMMDSSDIT--SSSEYEELDINDIYVPGSKYTLIEVLE 3410
 Db 2950 TPTLKEIVTDEGYDHY-----SDIYSSSEYEELDINDIYHVLINLTLEVLE 3003
 QY 3411 PSKR-----DIPSD--DTPSN 3424
 Db 3004 PSGKLSGNITPTSGKNTPSD 3023

RESULT 4
 T28432
 variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
 N:Alternate names: erythrocyte membrane binding protein 1 (Empr1)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28432
 R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinel, F.; Herfeldt, J.A.; Peterson, D.S.
 Title: The large diverse gene family var encodes proteins involved in cytoadherence an
 Reference number: 220487; MUID:95330813; PMID:7606788
 A:Accession: T28432
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3078 <SNX>
 A:Cross-references: EMBL:L40608; NID:9886374; PID:9886375; PIDN:AAA73396.1
 C:Genetics:
 A:Gene: var-1
 A:introns: 2611/3

Query Match 18.0%; Score 3489; DB 2; Length 3078;
 Best Local Similarity 26.7%; Pred. No. 7.3e-155;
 Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

Db 109 NIKKSDQGDNDGACAPRYRLHVCQDNLQIEPIKTTHTNHLVDCMAKEGOSITOD 168
 QY 187 HP-HKGT-----SDACTALASFPADIGIVRGIDMKPVNVD-----KYETGLREVFKKI 235
 Db 169 YPKYQATYGDSPSOICTMLARSPADIGILVRGRDYLGNPQELKQROJLNNLTKTIFGKI 228
 QY 236 HD-----GMEYKNDVNDPOGSGNYIKLREAMVNVNRRKVEAITCDASYSGYFMOSSEST 292
 Db 229 YELKNGAEARYND--PE-----FFKLREDMWPNRRKTYKATTCNA--WNTYF--HATCNR 280
 QY 293 PLFSNPKCGHKQGVPTNLDYVPOYLRFWDEGEERCRNRKIKKLVKDSCH-----NDKE 348
 Db 281 GERTKGYCCANDQVPTTYDIPYQIHLRFEBNAEDCRKNNKIKDYKKNCKGKREKED 340
 QY 349 RLCSHNGHDCSTTITWKKILHLDNKCTDCSTKCYFEVWLGNOQAEKREKYEKIO 408
 Db 341 R-YCSRNGYDCEKTKAIGLRYGKQICISLCAQNPYVWMINNQKQFQKQKKVDEIK 399
 QY 409 SY-----LSNDKPFVNNINSE--YKQFYELKETOYAINDPFLNLNBSGYC----- 454
 Db 400 KYENGASGGSROKRDAGCTTTNYDYEKKFYDELKSEYRTVDKLEKLSNEICTKYK 459
 QY 455 --KGLPGEKDIPTNSA-----DDKGLFYREYCYQCPQCYGCDGIXYTH--KSDND 504
 Db 460 DEEGGTIDFKNVNSDSTSGASGTNWSQGTFRSKYQCPYCGYK-----KVNNGSSNE 515
 QY 505 RERVVN-----EDYKPPGVKPTNITVLYSGNEQDIOXKLEFNCN-----SS 547
 Db 516 WEKNNKNGKSKGLYPRKDEKGTITLISKGHNDIEKLKPKDEKNGQJINSNGSG 575
 QY 548 IYKDKNN-----QKWCYTKDENINRCKLEQNTENNDPK-----II----- 586
 Db 576 TGGSGGSGNSGROELYEEMKC--YKEDVYVYGVHDEDEDEYEVNKNAGSLICILNOKNKE 634
 QY 587 -----SEHFEELWYVYLLRDTIKAMDIKTKCI--NNTTHC--IDECNRNC 629
 Db 635 EGGNTSEKPELDIOKTFFNFFYVVAHMLKDSIHMKKLQRCIQONNRIKCGNNKCNDC 694
 QY 630 LCFDRVVKQKEEWMNSIKKLFTRKKNIQ-----OSYYSNINLFEYGF-- 672
 Db 695 ECFKRWITQKDEMGKIVGHF--KTQNIKRGSGSDNTAELIPDHQVLYLNQIEELKGD 753
 QY 673 -----FKYMDKIDDEAKMKELMENIKRKNPFSLNNRD-----YLENATELL 718
 Db 754 SEDASEKSENSIDAEA-----ELKHLREITSEEDNNOEASVGYTEQKINMDKLL 807
 QY 719 DILKETATIC-----KDNNTNEACEFTHNATTPCYKPRGCTG--PRKTIKELAO 766
 Db 808 NYEKDADLCLEHDEEKEKEGDNCEIEGENFRYPCSGESGNKRYPLANKVAYAO 867
 QY 767 YKRSAYEBARNRGLIKLKAHEGILYKRGRRKDKDMICRIMIKSNRNLFSGNPGCD 826
 Db 868 MHKAKTOLASRAGSALGDSLSAQFKNGRNGSLTAKGICIKINENYSDNSGNSGPGCT 927
 QY 827 GKGTGCGIOTRRVYGTIEWEDREHMKKDEHYIMPRRHICTSNLEHLOTDPHLYNGI 886
 Db 928 GR-DGDHGGVPRMIGTIEWENISGKKQFSYKNVFLPRRHEMCTSNLENDV-----GSV 980
 QY 887 V-DGLVNSFLDVLISAYEANKIIRMYKREKNNIKGPEVDP--KQOTTICARIRS 942
 Db 981 TNKDKASHLSLGVOLAAATDAEITIKRYKDQNNI---QLTDPLOQKQDEAKCRVARS 1036
 QY 943 PADIGIIGRDLMERNGDMVLCLOHLEFVFNHSLKG--KGNDKYND--APRYLL 998
 Db 1037 PADLGIINGRDMWMDKSDTMEITRLIVFNKIKKHGINDPNRYTDESCKRPAYKYL 1096
 QY 999 RENWMEANPAKYWEAMKCDIKYLKDKSGHOSSTOSYCGSDHTPLDDYIPQALRMATEA 1058
 Db 1097 RADWMEANHQAWEAMK-----ATKGIIC--PGAPVDYDIPORLRWATEA 1141
 QY 1059 EWCYKQKKEYDLKKECKEDKDKNGOGCTKESGCTKCEACGEYDIIGLWMEQNN 1118

Db 1142 BWYCKASOBYEDKIKKICADOMSGDKG--CT--OGDVDCGKCAACDYKKEIEKWEOWR 11199
 QY 1119 IISDRYKELHDOQOMSVNSGSIENSSTAKNHID--RNVLEFSELYOONGK-----SNK 1171
 Db 1200 KISDKYMLIQAKTSTNG-----RVLGDDDDPDYQOAMDFLPHKASIAARVLVKRA 1255
 QY 1172 SGVSDSEAVIGTWYENVAAYLHDGNFDCOSONEFCEK-----SDGKUNEYAPR 1225
 Db 1256 AGSPTEIAAAPTTPYSTAGYIHOEIGYGCQEOOTOFCEKKGANSTSTTKENKEYTEK 1315
 QY 1226 DKPODHDGACGSGSGSPRTVOJTKTKKKAKEKTEKTYNDILKENDKQOVEDCPKRN 1285
 Db 1316 QPREPVALADCDINRSOTE-----EPKKKEENESACKIYEKILEGNGRTYGEBCPKES 1371
 QY 1286 SNGYPRMOC--GNTINLVEDPVCMPRRORICVHFLANDNEIKLOSQVNLKEAFIKSAA 1344
 Db 1372 ---YPPMDCKNMIDISHD--GACMPRRORICLYIHAESGTEINKTDMLKOFITAA 1427
 QY 1345 EPEFWSYXXK--DGBGNLDELKELKRGKIPPAFLBSMFYFGDYRDLFGTIDISKHGE 1403
 Db 1428 EYFLSQYTKSKNDSEAKTIDR---GLIPSOFLRSMYTFGDRDCLNTDIDISKQNDY 1483
 QY 1404 SKLEQIDSLFKNGDQKSPNGKTOEMWTEHSEHEWEAMICALVIGAKKD--FTENY 1460
 Db 1484 AKAKDKIGFEFSMDGSKSPSGLSROEMWKTNGEIKMGICALIKYVDTDMKRTKNDY 1543
 QY 1461 GYNNVYFSK--STLLEFKAPOFLWLTETWYDDCYTTOKYKLDQOENKCS--NDOLKCD 1518
 Db 1544 STDKNQSONGSPSELEFAKPOFLMWMLEMGEEFAEOKENIKIDACNEINSTQCN 1603
 QY 1519 T---BCKNCEDYVYKYM--KKEWLPQDXYRDERDKKRFDOHIGVMTDYGGINADY 1574
 Db 1604 DAKHGNQACRAVOYEVEKKEFGSGQTNNFLKANVOQODEYKYEYKD----- 1654
 QY 1575 IAKRTTASCGDKPSSAVQONILIEQOAYYADAKHCGCTFIENDKYTNISSKDKC 1634
 Db 1655 ---GVOP---IOGNEYLOK---CDNNK--CSC----- 1676
 QY 1635 GLVKEANTGAIKQKNGPNNNNKLEFEDVLFPSRLRICHALDNGTYDEYDENG 1694
 Db 1677 --- 1689
 QY 1695 KRRLMEVATEGYNLGOYKKEKEKEKIKTSDAHKSYEVPSCAMKYSFYDLNDIIGI 1754
 Db 1690 ---PRGXT---AHKYP----- 1699
 QY 1755 DNLDEKQTEBNLKKITNKNGTSYGKSDSTGTGNGSTARKFFWENKECWNNAMICGY 1814
 Db 1700 --- 1699
 QY 1815 KRGRDNGSGNSARDEDLKKGSGVSPDDDYPMGKNRDEGTAYOFLRWFAEMGEDFC 1874
 Db 1700 --- 1699
 QY 1875 EKELEKLVGACNDYTCGDNEDKRRKCTDACYQYKKEFISEMPPQYKQIKKYGENKDKIYS 1934
 Db 1700 ---EKC----- 1702
 QY 1935 EHPYAKADEAREYLDKQKIKCENKSDCEYKCMKDYSTORLTDGNSQNNPASPDLDEPK 1994
 Db 1703 ---DC----- 1704
 QY 1995 EVEGCMCOVPRGPVRRRETSPRVSLSISKATSKKEAKTAPPTKQPKVENITTEMRA 2054
 Db 1705 -YGCK--HVPSTP-----PPPVOPQ----- 1723
 QY 2055 QTRRRRAAOQTRKRTSTATTSTESDVGMVAKAISNNKPDSCGIEGCMPTKYGY--PKWG 2112
 Db 1724 ---PEAPYTYDVCSIVYKLEK--DTNNFSDAGLK--YGRKAPBSWK 1764
 QY 2113 CI-----VGKSENGNGICMPRRKIKCINNIOVLNET-----EKRD 2151
 Db 1765 CIPSDYTSAGAGATTGKSGSDSGSICIPRRRRRLYVGLQ--EWATALPOGEAAPSRSRA 1822

QY 2152 NDIKEAFICAIETOFMLKTIIE--NPAENE-----LQNGT 2188
 Db 1823 DULRNFIOSAIEFFFLMDRYKEKKPOGDSQOALSQLTSTYSDEDEPPDKLLONGK 1882
 QY 2189 IPDEFKRIYTYGYKDMFEGTDSNDKKITVNSYTTILNENKKOKKDEK 2248
 Db 1883 IPPDLRLMEYTLGIDYRDLVHGNTSDSGNTGNSNNNIVLEASGNKEDQKIOEKLEO 1942
 QY 2249 I-----FWENKKEFIWEGMITYLT-----HLADENEKIR 2280
 Db 1943 ILPKNGPILPVKSSAQTPDKWNEHAESIMGMICALTYEKNEDTSARDENKIEDD 2002
 QY 2281 DNY-----QYN--DMTKL-----TPSLEFYKRP 2302
 Db 2003 EYIEKFFSGTADKHGTASTPTGYTYKYDYREKVLKEDTSAGAKTSASDFTLLSDFLVR 2062
 QY 2303 QFLRMFTEMAEFECCKRREOLIKLEAGCKEY-----CN--GSN 2339
 Db 2063 PYFRLIEMGONFCCKRRKHLAQIKHECKVEENGSGSRGITROYSGDGACEMLPKN 2122
 QY 2340 DKTQF-----CAEACVYQNFIKKWTYEROREKFKKD--GKKYADYSTERDI 2390
 Db 2123 DGYVDLEKPSCAKPCSSYRWKIESKKEFEKOKRAYPOKDCVNGSNKHD----- 2174
 QY 2391 EKATCAHEYLMKLIKELGCKNKGCMQKPSQILKTTQOOSSPANDMPESLDYVPEEPN 2450
 Db 2175 ---KFCETL-----TSSAKDFLTKG----- 2195
 QY 2451 KCEPELSKSGSMI-----HTKKTIP--KIPMNCVEKAAAYLSKEANNMDITLKE 2500
 Db 2196 --PKPNVNEGKTIIPDDDKTEPKTKD--CDPCLKPSVNC----- 2230
 QY 2501 KFIPIESTKEKESKNSMTNNNPPCKPKPYADKTIGRNCPENFEENRFVYEWKCYKN 2560
 Db 2231 ---KQECNS--KGTDCRN-----KN 2247
 QY 2561 SKFYOEKRRVCVPRRHHMCLINLDEIKIRLKDNYLLKMWRTARNEGIDIIKNFNS-- 2619
 Db 2248 S-----IDATDIENGVDSTYLEMRVSDSKS-----GFNGD 2278
 QY 2630 ---ENGCMNPICDTMKYSFADLDGYRGTMRLIGYLPVEIKLYKVEYIYGKWRKN 2677
 Db 2279 GLENAC-----RGAG-----IFGCI----- 2293
 QY 2678 KGRKYNDVQTFERSAMWANDANKDITKA-----MTCAPEDAKLFRKGRDGEFRTILLO 2731
 Db 2294 ---KDEMKCRNVGYVYCK--PENVNGEAKG----- 2321
 QY 2732 DKCGHKDDPPVDYIIPQRFWMTEWSEYCYKALMELEKFKSCDHQKTSDRCKNDYDEN 2791
 Db 2322 ---HIIQIRALVYRWEIF--FEDYKTKH-----KISHRIKN--G 2355
 QY 2792 KCBOCKTRCOEYENFYLKWSLFDIOSNNKYLEEOPITYKISTYDHYONFVOKLTFKS 2851
 Db 2356 EISPC-----IKNCVEKW--VDOKRKEWKEITER-----FRD 2385
 QY 2852 ECVSESEYELHETSKLNYKFNENDGSSNIRYAREETPK-----SYKRA 2897
 Db 2386 Q-----YK--NDNSDDNVNYSFLETLIPQITDANAKKVIKLSFGNS 2426
 QY 2898 CSGTILSKNPLDNCPLDQNDKCGKELQTEFTPCSKNDYDNNLDMWNAVLYLVNSDDNKGLV 2957
 Db 2427 CGGSASA-----NEONKNG----- 2440
 QY 2958 IPRRRHLCRPIITAVANYKKQKEILKLLKLLSABSOGOLLOKYKSEBELCFEAMKYSY 3017
 Db 2441 --- 2440
 QY 3018 ADYSIIKGTDMQDLSSEKIKKIFETSNEATENKRTWENNRRQIWMHMLCGYKIATSK 3077
 Db 2441 -EYKDAL--DCLLKLKDKI----- 2457


```

Db 887 GTPSG-----KDT-----G 895
Qy 1296 NINIVEDPVCMPRRKOLCVHPL-----ANDN-----1323
Db 896 SI-----CVPRRRKLVGKLIHMAAGETTEAKSOETSGOQTSPSGNESSPSEKLPQ 947
Qy 1324 ----EIKLOSQVNLKAEATKSAAEETFSWYKSK-----DGEg 1360
Db 948 GHPPEETKEPESSELHAFAVSPPLRLRPLRMHFKPEQWKAHGAGATGOOTIITIGLDGG 1007
Qy 1361 NEL-DKELKEGKIPPAFLBEMETTFGDRPLFG-IDI---SKHGEGSKLKEQIDS---1412
Db 1008 EEPDKLKTGHTIPPEDLROMFTYLDGNDIIVHTSGNEDMOIMEAIOKIE 1067
Qy 1413 --LEKNGDOKSPNGKT-----ROEMWTEHSEIWEAMLCALV-----1447
Db 1068 OILPTSSSPSPRVTOQSIVENPKRTWNNENGGKIKWEGVICALIYNTDTPSGTAPTOI 1127
1448 -KITGAR-KDFTEN-----GYNNVKFSD-----KSTLEEFARPOFLR 1485
1128 QEVRTLRBENSNNPKIPOYKYDOVKLDTSDAKTTGSPVSGEKLTPLDLFSRPYFR 1187
Qy 1486 WLEWYDDCYCYTROKYLKDOEKCK-----SNDOLKC---DRECKKK-----CEDY 1528
Db 1186 YLEBMEGTCKEKKKLEKIEKCEKRDRTGHEHCSDGYDCTRTDADRNDKFDVILNCRCO 1247
Qy 1529 VKYMKKKEM--TPODKYYKDEBKKRFRDROHIGVAVTDTGTNATDILNRKFTASCGDK 1586
Db 1248 HICRCRYKRWIDIKFDEYHKOEK-----KYOG---EY-----DK 1278
Qy 1587 PGSASVVOQNIOLLEKOAYADAKHGGCTKFTIENDDKYNNISKDKCKGLVEANTGATK 1646
Db 1279 -----LNDKSSGGDNMC-----CKDIEKHKSAVP- 1304
Qy 1647 MONKGNYNINLKELEDEVLFPSRLRLICFHALDNGYTDPEYKDENGKLRLMEVAATG 1706
Db 1305 -----LKEK-----HCKNGTSENKNGOEDOL-----1327
Qy 1707 YNLGOYKKEKKEKEK-KTSDAHKSYEVPPCSAMKSYFYDLRDIILAGIDNLEDEKOKTE 1765
Db 1328 -----NKLDPKIPOTFSPSTYCKACP-----VYGV-----1353
Qy 1766 ENLKIKFNKGTSGVSGSS--TTGNPGSTARKFPMNENKECVNANMC-----GYKR 1816
Db 1354 -----NONGKRRGRGNGCTTINNE-----PENKENDGAASITSLINDGSTN 1397
Qy 1817 GDDGNSGNSARSDDLKCKGSPSDDYPM--GKNRDEGTAYOFLRMFAWGEDFCRHK 1874
1398 GATNGTGTG--TDETLKEC-----SDKYAFKGLRQOEWTCOK-----KGVNQO---1440
Qy 1875 EKELEKLVGACNDYCGNEDKRRKCTDACYOKKFISEMKPOYKOKIKKIGENKDKIYS 1934
Db 1441 ----NLNRRVND-TYFKD-----IVNFEFQRMWLYF--VHADNILLKHKI--1479
Qy 1935 EHPVAKDAEDAREYLDKOLKICEKNSGDCYKCKMJDVSTORLTDGNSQNPASLDEPK 1994
Db 1480 -----DPOIKKEKQOKT--EKKCI-----NCCNI-----1501
Qy 1995 EYEGCNCQVPRGPRVRETPSPRVSLSIKATASKKEAKTAPPTOKPKVENLTTEKRA 2054
Db 1502 ----KCEC-----VR-----1507
Qy 2055 QUTRRAAOQTRKRTSTATTTESDVGVVKAALNSKPDSCGIGCNPKTYGQYPRKMGCI 2114
Db 1508 -----KMLEI 1512
Qy 2115 VGKSEKENGICMPRRKCLICINNIOYLYETENKRDNDIEAFIKCALETQFLMLKYI 2174
Db 1513 KG--NEMG-----NIK--KHYNINSDDOKET-----1534
Qy 2175 IENPAEHELONGTIPDEFKRIIMTYTGDKDM-FEGTDISNDKKIITVNSVTIINEN 2233
1535 -----IAYNKSTYVDQGLFTDIDYKKAQKV-----1560
Qy 2234 NKKROKKKDEELRKLTEWENKKNKFIWEGMILGYTHLJLDE-NEKKIRIDNYOYNDMTKLT 2292
Db 1561 -----BDEKERKKIW-----GCTGH--DECEKKEKEK--NITNLI 1594
Qy 2293 PSLSEFPKRPOLRWMTLEWAEFCNKRKQOLLKLEAGCKEYEGNSNGDKTOEGAEACVT 2352
Db 1595 SELDDKITSQO-----NKH-----NPKGT-----AC--1616
Qy 2353 YONFLKMKTEYEROREKPKKDKDKKYKDYPTERDIEKATCAHEYLNMKLKELCGKD 2412
Db 1617 -----1616
Qy 2413 CSCMKPSSQLPKTTQOSOSDANDMPSLDVVPBEFNKCEPDELKSKGSMIHTKKITEP 2472
Db 1617 -----DPPSPPEETDPLDDTDPDLDDQHTBOQKPCP-----P 1653
Qy 2473 KIPANCVEKAAVYLSKEAENNDITLKEK-----FPISTK-EKESK 2514
Db 1654 PPPMTCVEKIAKELVEABGAKINNELKGNKNGDFNGKCNVKKNGCAVIGESCKFEQTYE 1713
Qy 2515 NSWTNNNCPDKPKPYAPDKYIGRRNCPENREBREFKVDYEMKCYKNSKFYOE-KKRVQVP 2573
Db 1714 NSVNNIN-----NKKDONNERKTIQOKM-----NFKYGTITRKDLICIP 1752
Qy 2574 PRREHMCRLNDELKIEELKOSNYLLKAVRRRTARNEGIDITIKFNSENGCAMPIDCTMK 2633
Db 1753 PRREHMCIDLISMLGRTTISOSSALKKIOEAKSERDITIRKILEQNSCEHRIICAMK 1812
Qy 2634 YSFADLGIIVGTDMLRIGGYLPPYEIKLVFEYIYGKMN--KN-----KGNKNNDVQTF 2689
Db 1813 YSFADLGIIRGRDMKNNNSKOKGLQKLEAFINITYNKLNDNNKYEKDKPKYLOI--1869
Qy 2690 RSAMWANDRKRDIWAKTACKADEAKLFRKGRMD-----GEPRIYLIDOKSGHKDDPVD 2744
Db 1870 RSDWMDANRKHIMWAMCMNAPDADAKFLKKNPDNSGSSSSKGIWTHSGNGYKPEPPDYD 1929
Qy 2745 YIPORFWMTESRYCCALMEELKFKKSGDHC-KTSDCKNDYDANKBOCKTRCOEY 2803
Db 1930 YIPORFWMTESRYCCALMEELKFKKSGDHC-KTSDCKNDYDANKBOCKTRCOEY 2803
Qy 2804 KNYVLKMSLEFDIOSNRYKLEIYEPYI---TKISTDYHONFOYOKLTKTRSEC-SWESF 2858
Db 1990 KKLJHNMKLGHF---KYKEIYNE-IYNNKDSKINSNEYKFKLEK---KDKCKELNNS 2041
Qy 2859 SEYIHTSKCLNTYKFNNDGSSNIRTYAFEBETPRSYEACSCILPSPKNPLDNCPTDO--N 2916
Db 2042 DKCIDETHTCKTKYFS--NSEKKNNNNTAFKPNPREYERKACDAP--DPLDNCPRKSATY 2098
Qy 2917 KDGCKELOTFTFGSKNDYDNNLDMWNAVLYLNSDDKRGVILPPRRRHLCRPLITAVNR 2976
Db 2099 EKACNTLLPTRLCKSKTFNNDDSDWDSFVQTSBPDRTGYLVPPRRROLDLKNITTY-KLR 2157
Qy 2977 KGDK-ELKKKLLTSAFSQOLLGOKYKSEELCEFAKYSYADYSDILKGTMDMTSLS 3035
Db 2158 SIEKIDDFKALMTSAVNEKGLCELYKKDRDVTLOAKMKSIFYDGYIVAGTDLSTAPL 2217
Qy 3036 EKIKKIF-----ETSEATENRKTWNNRRROIWAMILCGYLIATSKYTLIDSGOULR 3089
Db 2218 DKLTKILNVLKLDGTYNEIREDKGMWTEKNTKRWAMILCGYLAAGKI--EERDCSLP- 2274
Qy 3090 DEETNOFLRWLIMWAQACEKRVKVSLSLTKCPRSNEDNFEASELLRPPGOO---NDIR 3146
Db 2275 DDWTHQFLRFRBMSSEHFCAROKLFNEYKREC--ASACIIIEYGITDPPVCEACTQYR 2332
Qy 3147 KYISLNLKNTYENNLIKYKOLKDOSSGNDKPSB-ENVOYSIKSKDSOCALDELNDIN 3205
Db 2333 DYITRKIQ-----EYRLNLYOYNTNFENNEKKAELVAKAEFYNDK---CNDKONGLS 2379
Qy 3206 EIYTGKKNNEENEFKVLKLYGLYFVEDETHKNHVNUNGINKEEQOYRPAALAFPTFH 3265
Db 2380 KYI-----DIEKKWMMYDSF---DDNLLKKNKCIROIKRKP---PKVAPREEH 2424

```

```

0Y      33 KSNVNLVERKAKNNIRHPSKYAKEHVUS-----LGGDLTKAFERQGRSPRPVKKHHYYYP    85
        17 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  :
Db      17 R$AHLLDLSICKY- ----TDKHGDADLPDSNGKLKTJSLAIFEKAPESGOTSSE- ---   67
        86 YPCNLDHKHEHNLE- YDVNVLRHPCHGROPNFEDEESEC- NKIRNYKRKNDAIACA  142
        68 ::::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  ||
Db      68 -PCLNHBYHTTVISGVDKEN---PCMDREVPFVSTEGMECCKSKIRGSNSKKD- ACA  122
        143 PPRRRHMCOKLEAL-NDINTONIHLLGNLVATAYKEGESI-----VNHPHKGT$  193
        123 PFRRLHLCDOHLEIKHIKDKITR--HNMLADVCFAKAFFEALESIEKYGOYLNN-----S  174
        194 DA-----CATAR$ADIGDIYRGIDMKRPN--VHDKVEICLR$EVFKIHGHMD-EVKND  246
        175 DVNINICEIAR$PADIGDIYRGDILRGDKDERIRKEENLRKI$FKKIDYNLDADAIOEH  234
        247 YNPD--SGSGNYKFIREAMMYNRNRKVAEATITCDASYSGYFMOSSESTPLFSNPKCGHKO  304
        235 YKDDDKGKGNKTKLRNMWMEANRQYVKAALTCCGA--GGTYFPOTGCTGWTEKRCPI  292
        305 GKVPNTLDYVFOYLRFMEDWEEFCRRRNRIKLKRVKDCSRNDKERLYCSHNGHCYT7TM  364
        293 NDVPTFYEDYPOYLRFWEEMAEDECRRKKKKYVIDKNCNCHNTSRNIYCSCNGDLCJETIR  352
        365 KKGILINDKTCOSTGCNVFEWGLNGOORAFKOKRKYEKELIOSYLSNDNKRVNNINISE  424
        353 VIGHNVIGSECSK$WCVRK$KKNIDNOKEEFLKOKKCENEMLS$SKK$OSTRYVNYEG  412

```

QY	425	YKQFYKLEKETOYATJNDTLNLLNGKXK	-----GGJPGKDDITFNS	469
Db	413	YDKFEFYKLLKSNVGGJDKRLLELLNBERCQFESNDLCKITDKYDGGGVALIGGASDS	-----	472
QY	470	ADDKGIFRSEYCOVCPDGCVCODGIKYTHKSDNREVRANNEDYKPRPVGKFTNTVLVS	529	
Db	473	NNSKRTSHSOYCEECGCCVELIGDEM - KEKNKECGCKGRYINPKGTJHNVPLVLSF	530	
QY	530	GNEBGDTOKLENCNSSTNYKDKNNQKWCYKXDNINRCLBQNTLENDNPKI - ISF	588	
Db	531	GDEKEIIEIKEDPCASNSDSKLBQMKCIYGDKEYEYCLTEJNRNKSSEEDPEEIQTFE	590	
QY	589	HNFELWYTLJLBDTJIKMNDKLTQCTINNTT - HCIDECNRNCLCEDRWYKQKEEEMNSIK	647	
Db	591	HNFYFWYRHLNANDSIEMPRKINNCIEKAEKCKNECKTDCGCFOWIRGKKEEMGEIK	650	
QY	648	KLFPKKNNIOQSYYSINNLFEGYFPRVMDLKDQDAKKKLEMEINIKRKNFESYLENR	707	
Db	651	KHEF-----TODGFSIFGN---NYDYLEVNLID-----ELFODITEAYNSOKIQIG	697	
QY	708	DYL-----ENAEITLDBLKETATICKDNNTNACESTSH - NATPNQVCPVR	752	
Db	698	DLTKKKTQAADATBEQKNITIDLLFEYDSEAKKC - KIOEGCPKPKTYKRNACY---	752	
QY	753	GGTOPYTKNIEIAOYFRRSAVEEA - RNRGLHKIKGAHGIYKRGRRKDFDNLCRMTI	811	
Db	753	GNNTYDALAGVAOKLOOEKKEQDLDRNDSKALKAASOGKSNQGDPPDEFKNNLCGTQ	812	
QY	812	KHSNRNLGFSRPGCDGSDGDIOTFRVYVTEWEVDPEHMRKDHEVIMPRRHICSN	871	
Db	813	KHSNA - IGDSNPNKNGK-----KERPNVGEKKNGGE - VKASHDYLPPRRHJCTSN	865	
QY	872	LEHLOTDPHPLNGNIYDVLNNSFLGDVLLSAVEANKITIRYKKEKNKLGRKETTDPKH	931	
Db	866	LEHINTK---STGLTSDKAHSHLSGLVLLAAK-----KEGEDJ - TKLTENDN	909	
QY	932	OTTICRAIRSFADIGIITLIRDLMBENGVMYLOHLETVYGNITHKSLKGNKDKNYD -	990	
Db	910	RSSICRTMKYSFADIGIINGTDLMDINGATGVONNLKDIYSKITEELKCKHPKPFEND	969	
QY	991	----DAPYTLKLRNMMWEANRAKWEAMKCDIYULDKDSGHOSTOSSYCGSDHTPLDY	1048	
Db	970	DKYINDSKHTRLRSMDWEANRDQWYKMCPTK-----NCGIQG---ATPHDY	1016	
QY	1047	IPQKLRMTAEAEYCYVQKKEFDKLEKCKECKDXDNOGQCTKESGTCYCTACBEY	1106	
Db	1017	IPORLRMVWEAEWFCEQSRLYEELRLDQSG---TTKQ - CNNDK---CAKCDQOCBEY	1063	
QY	1107	NDIIGLMEQWNTISDXYKLEHBOQMSVNSIEASSTAKHIDBNVIEFSELYOOG	1166	
Db	1070	KTKIQPADQONEISNKTQILYMQAKIAALNGSTEBSTTKDKDKKNVLDIQLKHEANY	1122	
QY	1167	GKSNNKSGTSDS-----AVIGTNTYVENGAVLHDTGNFDDCQSONEFCEKSDSKD	1218	
Db	1130	G---TRGPPRAHEDRRPRRAATSKSDVYETTAGYIHOEARTRECLQWVFCNNNG---	1182	
QY	1219	NEKVAFRKQPDHDHAGCGCKSGSPRVOYKTKKKKABEKDTECKYVNDLKENDDKQYE	1278	
Db	1183	NNETAFLSTLPEHYHAACK-----NNKKNASS-----PEELGRSDSF	1218	
QY	1279	DCHEKKNNSNGYPRMOCNINILVEDPRVCMPPROKQJCVFLANDNEIKKLOQVMILKEAF	1338	
Db	1219	DDH-----QIPREDEEVIH-----	1233	
QY	1339	IKSAAEFTFSWYTKSKDQGBNLEKLEKGIIPAFRLSMFYTFGDIRFLPCTDIK	1396	
Db	1233	-----SSEEEDEESEDDEEKE-----	1248	
QY	1399	GHGEGSKLKEQIDSLFKNGOKSPNGKTRQEMWTBHSHEIWEAMICALVKAQAKDDPTE	1456	
Db	1249	-----BEVEEYHNGADEKA-----	1262	
QY	1459	NYGINNVKFSDKSTTLEEFAKROFLRWLTLEWTDYCYTRQRYKLYADVOEKCKSDNLKCD	1518	

```

Db 1263 ----- 1262
QY 1519 TECNKKCEDEVYKMKKKKEMIPQDKYKDEBDKKRPRQHGIVNTDTYGTGNATDYLRK 1578
Db 1263 ----- 1262
QY 1579 FTASGDKPGSASVYQNRITOLEKQAYYDADKHCCTKFTENDKYNISSKDKCGLVK 1638
Db 1263 ----- 1262
QY 1639 EANTGALKMOKGPNNTNKLKEDLYLFSRRLRICFHALDGNITDPEVKENGELKRL 1698
Db 1263 ----- 1278
QY 1699 MEVAATEGYNLGOYKKEKKEKIKTSADAKYSEYVPCSGAMKYSFYDLIDILGIDNLE 1758
Db 1279 ----- 1296
QY 1759 DEKQTEENLKIFNKNMGTSVGKSDSTGTGNPGSTAKFEWENKECWANAMICGYRGR 1818
1297 ----- 1296
QY 1819 DDCNSGNSARSBDLKKCGSVPSDDYPMGKNRDEGTAYQFLWFAWEGEDFCKHKEKL 1878
Db 1297 ----- 1308
QY 1879 EKLVGACNDYTCGDNEDKRRKCTDCTOYKKTISEMKPOYKQIKRYGENKDKIYSEHPV 1938
1309 ----- 1310
QY 1939 AKDAEDAREYLDKOLKIKCENKSGDEYKCMKOVSTQRLTDGNSQMKPRLSDEPEVEG 1998
Db 1311 ----- 1328
QY 1999 KCQCOYPRPPRYRRETPSPRVLISKATASKREAKTAPPTPOPKVENITTEMRAQRT 2058
Db 1329 ----- 1364
QY 2059 RRAAOQTRKRTSATTTESDVGTMVAILSNRPDSGSGEGCNPKTYGOYPRMGCTVGKS 2118
1365 ----- 1395
QY 2119 KENENGICMPPRKRLCINNIOYLYETENKRNNDIKFAIFCAIETQFLKTYIE 2176
Db 1396 KESSSG-----GKEPPSDK-----LRAFIESAVERPFLMDRKKKML 1434
2177 ----- 2213
QY 1435 AOKKAELOGLDLYSSGDGDPDPN-----ONKILNGVIPPDLFLAMEYTLGDYDLIVHGHN 1492
2214 SNDKITTNTSVTTLNENNNKODKKDELRKI-----FWEK 2253
Db 1493 TSDSGNTGNSNNNIVLEASGNKEDMKIOETIEDILKNGSTPLVPKSSAQTPDKMNE 1552
QY 2254 NKKFTMEGIYLYT-----HLTDNEKEKIRDN----- 2283
1553 HAESLWKGICALIYTERKNPDTISARGDENKIEKDEVEYKEFGSTADKHTASTPTGYK 1612
QY 2284 -QYN-DATKL-----TPSLEFYKRPOLFRMTEVAHECNKRKQDLKLE 2327
Db 1613 TOYDEKVKLEDTSGAKTPSASSTPPLISDYLVRPFRYRLEEWGQNCCKRKKRLQIK 1672
QY 2328 AGC-----KEYECNG-----SNDKTOECACACTYONFLKWKTEYEROREK 2370
Db 1673 EECMDGSDKYSYSGDEQCDRBDTSNEVSADLEGRSCGNSCFYKWKIKRKKKEYDKQANA 1732
QY 2371 FKQKDGKKYDYSTEDIEKATCAHEYILMKLKELOGNKDCSCMOKPSSQLPKTTQOS 2430
1733 YSKOK-----TKYF-----EGS 1744
QY 2431 QSSDANDPESLIDYVEEFNKCCEPELSKSGSMHTKITEPKTIPMNCVEKAAV----- 2485

```

```

Db 1745 KGAGLND-----HNKECYKLG-----TCTDAAFILRLKN 1775
QY 2486 --LSKEAEN-NDITL-----KEKIPLESTYKESKSNSTNNPCDPKRYADKTYIGRN 2539
Db 1776 GPCKKDNENGNDINGNTIEETFRPAENCK-----PCSSFK-----INCRN 1816
QY 2540 -PCENNEERKAYDYMKCYKSKFYQEKKRYCVPPRREHMLRLDEIKIERLDSNYL 2598
Db 1817 GNCRSGDG-----DYKEKC----- 1830
QY 2599 LKAVRTAENEGIDILKNFENSEGAMPICDPMKYSFADJGDIRGDMIRIGSYLPPV 2658
Db 1831 -----NGGITTTGNNTMGCT-----EDVV----- 1851
QY 2659 EIKLKVFEYIYKMKRNKKNKKNKYNDVQTFRSAMWANDRKDIWAKMTCKAPEDAKLPK 2718
Db 1852 -----MHVSDKNMNEFEGDGLDACEBAGIFGTGRKDEMK-----CRVYCGHLICQ 1898
QY 2719 GRMDF-----ERTLLQDCGKHDDPPVDYIPQRMMTMSSEYCYKALMELEKFKSC 2775
Db 1899 EKGNGAINDQITLVR-----ALIKRWEYF-----LEDYKIKKKL 1935
QY 2776 DICKTSRCKNDYDENKCEQCTRCQYKRNFLKWKSLFDIQSNNKYEY----- 2825
Db 1936 KPC-----IENGNGSTCINGCNKKNRYGEMIKLKQEMWIKINHYLEKKNKEDKN 1986
QY 2826 -----EOPLYTKISYVD-----HYQNFYQIKLT-----FKSCSVSESEYIHTSKCLNRYN 2874
Db 1987 VTSLATVLETLVTOILAANKRREOTSLDKLTSLGNCNCPNS-----R 2030
QY 2875 ENDGSSNIRTYAFETPRKSYEACSTLPSKNPLDNCPTDQNKQCKELOTFTFCSKNDY 2934
Db 2031 KNDGNE-----DAIDCM-----NKLEKIKHECTQH 2058
QY 2935 DNNLDNNMAYLVNSSD-----NKGYLPPRRRLCTPRITAYNRYKQDKELKLLTSA 2991
Db 2059 ENSVEN-----SDQPHNCGGNPP----- 2078
QY 2992 FSGQLLGQKYKSEELCFEAMKYSADYSDIINKTDMTSLSEKIKKIFENSNEATN 3051
Db 2079 -----DEEDLLER----- 2087
QY 3052 RKTWENNRROIWAMLGKYKIASVYTLDEGWCQLPKDETNQFLWLEMAKQACKER 3111
Db 2088 -----ENPVEQ-----PGCFPTQOQBEPD----- 2107
QY 3112 KHVDSLTKCPRSNEDNFEASELLRQPGQONDIRKYSILNLIKNTMENLITKYQLKD 3171
Db 2108 -----DKCGKLEKKDEKE-----OP----- 2124
QY 3172 QSSGNDNKPSEENVQSYIKSDSOCALIELNDINEIVTGTKNNEBNEFKVLEKLYPGY 3231
Db 2125 -----EOPAEEDGALIVPSG-----PG----- 2142
QY 3232 FVEDETHKNVLDGNIKKEEQYRPAKALYFTPHVDSFYQALPFSYTHRAQDIPKNDILK 3291
Db 2143 -SEPEADKGVYKAEL-----PKP-----QEPDLSHPA-----VIP 2173
QY 3292 SSISVIVTSLG-----IALHFKKKKSSV-DLRLILNPOGEYMPLESKNRYIPR 3346
Db 2174 SLVSTLWMSVIGIGFAFYFLKKTSSVGNLPIQIDPKSDYIPIPLKSPNRITPT 2233
QY 3347 SGPKYKGTIYMGDT-----SGDEKTMMDLSSSDITSSSEYEELDINDIYVPGSKYKT 3403
Db 2234 SGKTRGRIYTLGDSGTDSGYDHY-----SDLTSSSEYEEMDINDIYVPGSKYKT 2287
QY 3404 LIEVLEPSKRDIPSD-----DTPS-----NOTPRKNRTIDEMNEL 3440
Db 2288 LIEVLEPSPGNNTTASGNNTTASGNNTTASGNTPBDQNDIQNGIPSSKITDEMNOL 2347
QY 3441 KHDVSOYL-----PNTPENNNYSADIPMTEPNTLYSDNPEKPLIISHTDILTGKEI 3497
Db 2348 KOFESIOYLQSEBNTPE-----NMLGYVNDNNTHTPTSH-HHVEKPTIMSHDRNLSGSEY 2404

```

QY 3498 SY-----NINMSTN-----TNNIDIPNARNDSYRGIDILINDSLVLYMLLY--M 3539
Db 2405 NYDMFNSGNPNINISDSTNSMDSLTNSNHSYPYNDKNDLYSGIDILINDALSGNHIDYDEM 2464
QY 3540 MK 3541
Db 2465 LK 2466

RESULT 7
T28161
hypothetical protein FC3-var11.1 - malaria parasite (Plasmodium falciparum) (fragments
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28161
R:Herandez-Rivas, R.; Mattei, D.; Sterkers, Y.; Peterson, D.S.; Wellens, T.E.; Scherf,
Mol. Cell. Biol. 17, 604-611, 1997
A:Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.
Reference number: Z20483; MIMD:97154495; PMID:9001213
Accession: T28161
Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2647 <HER>
A:Cross-references: EMBL:067959; NID:g1794255; PID:g1809295; PIDN:AAC47438.1
A:Experimental source: strain FCQ27/PNG
C:Genetics:
A:Introns: 2158/3
A:Note: FC3-var11.1

Query Match 13.8%; Score 2686.5; DB 2; Length 2647;
Best Local Similarity 23.6%; Pred. No. 2.1e-117;
Matches 889; Conservative 379; Mismatches 861; Indels 1635; Gaps 120;

QY 31 SHSASNVLERYAKNIRHPSKYAKEHVS-----LKGLTKAEFRG-----PSEPVNK 79
Db 9 TNSAKAEVLDEIGETIQ-----KKAHSDAPFRSOLKNGNGEAKFYNGEIMQPNGL-- 61
QY 80 HNTYYPCLNDHKEHTNLRDYDNLHPCHGRENDEDESEC-GNKIRNYKRKND 138
Db 62 -----CELDITIDTINVDGHSN--PCEGQTVAFPPDNRSOCKKNRKKD-SVDNSV 109
QY 139 IACAPRRHRMCDKNLEALDIN---TONHDLGNVLTAKEGESIVNHH--HKGT 192
Db 110 GAQAPRRHLGCSHNLSEIQTNNYDSSKAKHNLAEVCYAKKEGESIVNYBQLGHHTT 169
QY 193 SDACTALARSFADIGDIVRGIDMEKPNVHD-----KVETGLREVEFKIH----- 236
Db 170 EGCICTLARSFADIGDIVRGIDMEKPNVHD-----KVETGLREVEFKIH----- 229
QY 237 DGMEDVKNNDYDGSNGYKYLEAMWNNVRNKNVWEAITCDASYKSGYFMSQESNTPLFS 296
Db 230 NGAEEERYK-----DGSNGYKYLEAMWNNVRNKNVWEAITCDASYKSGYFMSQESNTPLFS 284
QY 297 NPGCHGKQGVNPNIDYVPPYLWPFDEMGEEPRCKRNIKKYKDSR---NKEKRLYCS 353
Db 285 GEEGQCIDGVPNIDYVPPYLWPFDEMGEEPRCKRNIKKYKDSR---NKEKRLYCS 344
QY 354 HNGHDCSTTITMKGLIHLIDKCTDCSTCKVFEVWLGNQOEAFFKOKREKEKEIOSYLSN 413
Db 345 RNYDCTKTIRSIDKYSMNECTKCLYCDPYKWKIDNKKKEFEKOKKKEENELRYRNES 404
QY 414 DNKFNININSEYKQFYEKLKETQYATNDTFLNLNEGKCKGGLPGEKDTITFTNSADK 473
Db 405 SQNSPKRYNNMYETDEFGNLLKDYOSMND--FLKILNSETPCTNIDAKSKIDFTK--DPE 461
QY 474 GIFYRSRYCOVCPDGCK--CDGIKYTHKSDNREARNNDYKPRPWYKFTNTIVLSCN 531
Db 462 ETSSTHETCPCEPCMGCKTADG--TWKRLYENDPQCPKRYEPRKVELETEDIVLTGK 520
QY 532 EGGDITOKLENFENSSYNYKDNKNNQKWEYKXKNDINRCKLEONTEINNDPKIISFHN 591
Db 521 ENNDIILKLEPFCKTGNNTGFK--NEEMWCYYGYGN--DKCYLNGEELGGEK--KVKQVDNF 577

QY 592 FELMAYTLRLDTIKWMDKLTCTINNFTTTCIDECNRNCLCFDRWVYOKEEEMSIKKLFT 651
Db 578 LMFVVAHMLKDSIEMSKLSNCLSKDKKTCITTCNPNQCYDQWIGKVKHWTQIKKHF 637
QY 652 KKKNIQ-QSYSNINNNLFEQ-YEFKWKMDKDKDEAKMKELMKNIKRKNFSNLNNDY 709
Db 638 KQTFQGWCHFYVLEYVLEBDOFTTTRKAYGARELVHIOEMLOKKKEQV--LHDSAN 695
QY 710 LENAIELLDHLEKETATIC-----KDNV-----TNEACETSHNATTPCYKPRGTOPTK 759
Db 696 MKTILIDELDLHELEKACQIVNHNKDNKCPADLSDESEEDIDIQONKCKAP--SGNHRA 754
QY 760 NIKELAQYFRSAYEAPARNGL--HKLKKAHEGITYKGRKPKDNLCKIMIKHSNRL 818
Db 755 LVNKAASNMMHKKRROLVNRGVSSKLGDAAKEGRSGTIIKLD--ICSIDHSAKAR 813
QY 819 GFSGNCGDGGTGDTOTRRVVGTEWPEVDEHNRKDHEDYIMPRRHITSNLEHLOT 878
Db 814 GHIDQPCRRKDSKVYAKNR-----RWDVAGFISNTYKDIYMPRRHOFCTSNLEYLOT 868
QY 879 DHPNGNIYD--DIVNSFLGDLISAKYEANKIIRMYEKKNLKGPKRYTPDKHOTI 935
Db 869 NKLINGNDINGNINIINDSLGDLVLAANYEADPIKMYKQN-----DYKNATI 919
QY 936 CRAIRSFADIGDILNCR-----DLERNGDVYKLOGHLETVFGNIRHSLKG 982
Db 920 CRAMKYSFADIGDILQROHICRINIYERVKEHSEIRFLILSKNLT--AFKEI----- 971
QY 983 KGNDRYNDAPKYLKLENNWENANRAKVAEMACDIKYLKDKSGHSTOSSYGYSDHTP 1042
Db 972 -----KEDRP--YTKLREDWENARRKRIWEAMOC-----PPNGSFPCSKSHIG 1014
QY 1043 LDDYIPQKLRMTAEWYCYVOKEEYDKLEKCEKDKONGOGCKESGCTGCTKTEA 1102
Db 1015 LDDYIPQKLRMTAEWYCYVOKEEYDKLEKCEKDKONGOGCKESGCTGCTKTEA 1069
QY 1103 CNEYNDITGL--WKEDWNIISDKYELHQAQMSVNSGIEASSPAKHIDRNVIEFLS 1159
Db 1070 CKHVAKIKNLLIHGKEDWMEIRKYLLOAOTTAANGSPDYSGLVDENERPVYVFLF 1129
QY 1160 ELYOONGK-----SNKSGTSDSAVIGNTYENYCAVLH--DTGNFDDCOQSONEF 1209
Db 1130 ELYKENGKIGNPDIPRARRSKRETAPASVANKDYSTAGVYHDMGPHMECKTQTER 1189
QY 1210 CDEKSDGKNEKYAFRDKPDHDGACGCSGSKSPTRVQITKK-----KAEKD 1258
Db 1190 C-EKTDQYVNTTFRKPPQYDADACINTRPPKEDSKRSDESDSEEVYKETEYBEKA 1248
QY 1259 TE-----CTYVNDILKENDGKQVBDCHPKKNSNGP--DMOC-- 1294
Db 1249 TEDAVDTGPPAPKEATTTLDVCPYAGVL--KENLEMACPTKYGPAPKAPSMKCIPT 1304
QY 1295 -----GNINLVDPDR-----VCMPPRRQKLCV--HFLANDN 1323
Db 1305 EKTNAATGSESGSGGALQAPKATVYESGPTVNSGSLIIPRRRLLYLOKLHMASGN 1364
QY 1324 EIKKLOSQV-----NLKEAFIKSAAEFTFSYYSKSGEGNEJDELKEL 1367
Db 1365 TVVSGAQOTPOGGSPTSGKETPSDKIRTAFFISAALETFFLMDRYK--KKAIAKKE 1420
QY 1368 KEKQ-----IPAFLSMEYTFGDRFLFTGDT 1395
Db 1421 KKKOMVDYSLSTADPHNNVSLVIAPNPNYKTCVIPPPLFLOMEFTTLGDYADIFGKN 1480
QY 1396 -----ISKHGE-----GSKLKEQIDSLFKRGDQSPNGKTRQEMWTHSHIEAMICAL 1446
Db 1481 DIVIDTKNGDKDAIERKKTKIDAIERYLKNADSOQPRDEKROTWEONGSHINGMICAL 1540
QY 1447 VKIGAKKDDFTENYGVNNVFSKSTYLEEFARPOFLRWLTWYDYCYTRQKYLKDVQ 1506
Db 1541 T----- 1541

QY 1507 ECKSNQDLKCDTBCNKKCEDYAKYMKKKKEMIPQDYKYDERDKRPPRQHGIVAVTDY 1566
Db 1542 -----YK----- 1543
QY 1567 TGTNATDYLNKRTASCGDKPGSASVYVGNLQLEKQAYYADADKHCCTFIENDKYTN 1526
Db 1544 ----- 1550
QY 1627 ISSKDKCKGLVKEANTGAIKWONKPNNNYNNLKELTEDVLPSPRRLICPHALDGNATDP 1686
Db 1551 ----- 1550
QY 1687 EVKDENGLRKRLMEVATBEGYNIGQYKKEKEKIKTSAAHYSTEVPSCAMKSYFD 1746
Db 1551 PLKONBELKALMD-----EKNKKPKDQ----- 1577
QY 1747 LRDIIIGIDMLEDEKOTENLKKIFNNKTS-----VGKSDSTGPNSTARKPFMMEN 1802
Db 1578 -----DKYKLDEN-----SCTSPKIYVPAKPTTTPPPSPISF----- 1612
QY 1803 KECVWMAIGYKRGKDDGNSGNSANSDEDLKKGVSPPDDYPMGKNRDEGTAYOFLW 1862
Db 1613 -----SRP-----PYFRY 1620
QY 1863 FAEMGEDFCRKHKELEKIVGACNDYTCGDNEDKRRKCT-----DAGYQKKFISEMKPYE 1919
Db 1621 LEMMATFCEKRRKLEKI-----KVECHDEGKKQKQSGDGEDEIKR----- 1665
QY 1920 KQKTKGENKDKTIYSEHPAKDAEDAREYLDQKIKCENKSGDCEYKCMKDVSTORLND 1979
Db 1666 -----ODYSIVRDEYCEPGKCYR----- 1684
QY 1980 GNSQNPASIDEPKEVEGKQVPRGPRVAREPSPRVSLSKATSKAEAKTAPPT 2039
Db 1685 ----- 1684
QY 2040 KQPKYENLTTEKRAQTRRAAQOTRRKRTSTATTESDVGIMVAILSNKPDROGIRG 2099
Db 1685 ----- 1684
QY 2100 CNPKTYGOYPKMGCIYKSKENENGICMPRRKLCINNIOYLTENTENKRDNDIKEAFI 2159
Db 1685 -----FKRM-----IGKRRKDEYD-----KQKE-----AYNNOKTARNN----- 1715
QY 2160 KCAAIETQFLMLKYLIIENPAENELONGIIPDEFKIMYTYGDKMDFGDISNDKKI 2219
Db 1716 -----NDNAF 1720
QY 2220 IIVTNSVTIILN-----ENNNKKQDKKDE-ELRKIFWEKNNKFTMEGMITYGLVHLTDE 2273
Db 1721 STTLDTCTAGDFLOTLRKNGPCKNNDVDSGENKIF-----DE 1759
QY 2274 NEKERIRNOYONDMTKLPLSLEEVKRPOLRMTBWAEEPCNRKQOLKLEAGCKEY 2333
Db 1760 N-----GTFEKT-----OYCG-----TCSLNGF 1778
QY 2334 ECNGSNDKTOECACACYTYONFIKKMTEYEROREKFKDKDGKXYKDYDSTERDIEKA 2393
Db 1779 KONGND-----CAVRIN-----V 1791
QY 2394 TCAHFIYLMKLEKELGANDCSCMQPPSOLKRTQOOSOSPANDPESLDYVPEFNKE 2453
Db 1792 TC-----NGSNRTTITIAADI----- 1807
QY 2454 CPELSKSGSMHTFKRTEPKIPMNCVEKAAYVLSKEAENNN-----DITLKEKFIPIESTK 2509
Db 1808 -----KNG-----GNSAIIINLVSDDINSNGFNDLEA-- 1835
QY 2510 EKESKNSMTNNPCDPPKVPADYIGRRNCPENRENERK--VDYEMKCYKNSKFOEK 2567
Db 1836 -----CKN--ANIFGKIKENMKC-----YFCK 1857
QY 2568 KRVCPVPRREHMCRLNDEIKIERLKSNTYLKAVYRTARNEGIDITIKNFSENGCAMNP 2627

Db 1858 SDVC----- 1861
QY 2628 ICDTMYSPADIGDIVRGDMLRIGGILPVEIKLYVEIYKWRNKNKGRNNDYQ 2687
Db 1862 ----- 1869
QY 2688 TFRSAAWANDRKDINKAMTCAPBEDAKLEFRKGRMDGFERITLIDQCGHKDDPPVDYIP 2747
Db 1870 -----DONOILILIAL-----FKR----- 1883
QY 2748 OREFRWATESEYCCALMELEKFKKSCDCKTSPRCKNDYDENKCEQCTRCOEYKNFV 2807
Db 1884 -----WLEXF-----LDYKIRKRLNCP-----INGKALCTNGCVA----- 1917
QY 2808 LMKSLFDQSKKYLEIEOPIYTYKISTYDHOVQVOKLTFKSECSVESSEFYLHETSK 2867
Db 1918 -OW-----INHKRTWMTNLSF----- 1933
QY 2868 CLNYKFNENDSSNIRTYAFEEETPKSYKACSCILPKNPILDNCEPTDQNKDCKELOTF 2927
Db 1934 -----NEQYNGDDTERNRLRFFVDLIRQ-----IAATIDK-----GNHGLVLYKVS 1977
QY 2928 FCSKNDYDNNDNMNAVLYLVNSSDDNKGVLIPRRRHLCRPIYAYNYRKGDKEILKKL 2987
Db 1978 KC-----NCGN-----NSQNKBEENDIVLC 2000
QY 2988 LLSAFSOGOLGQKKSSEELCFEAMKSYADYSDIIGKTMODTSSEKIKKIFETSNE 3047
Db 2001 L-----OKLEKAKEC-----KNPETS-----GIPOOPE 2026
QY 3048 ATENKRTWENNRRQIWMALCGKIATSKYTLDEGWCQOLPKDEOTQFLWLEMAKQA 3107
Db 2027 VSPNH-----IEDE-----EOPLEEBN-----T 2045
QY 3108 CKEKHVDSLKTCQPSNENEFASELLROPGCONDRIKYSILNLIKTMENLNIKY 3167
Db 2046 VEHPRICDVLKHHNDROR-----LVKNPLVOPTLKRK 2081
QY 3168 QLKDOSSGINDKPSSEENVQSYIKSDQCALBELNDINEIYVTGKNNEEFKEVLKLY 3227
Db 2082 KKKKKRRKKIKK-----NODFH-----PRHL 2103
QY 3228 PGLYFVEDETHKNVLDGNIKE--EEOYVRKALYFTPHVDSFOAPLFTSRVAQYD 3285
Db 2104 PCGAFINTNTPKTKTPSSGKNPWEHFAVIPA----- 2135
QY 3286 KNDLKSSISVYIYVSAUGLIALHMKKKKSSVDLRLINIPQGEYGMPTLESKNRYIP 3345
Db 2136 ---LVSTLANSVGIGFAAFTYFLKRTKSTIDLTLST--IPKSDYDIPYKLSPNRYIPY 2191
QY 3346 RSGPYKRTIYIMEGDT---SGDEKYMMDLSSDIT--SESEYEELDINDIYVPSPKY 3401
Db 2192 TSGYRGKRYIYLEGDSGTSGYTDH-----SDITSSESEYEELDINDIYVPSPKY 2245
QY 3402 KTLILEVLEBS-----KRDIPSDTPPSNDTRPNTKFIIDENNEELK 3441
Db 2246 KTLILEVLEBSGKLSGNTIPTSGNNTYASDQNDIPTSDRPPIT-----DDEENITL 2298
QY 3442 HDEFSOYLPTNTEPN--NNYSADIIPMTEPNTLYSD-----NPEKRPITIS 3486
Db 2299 HDPLSNMLQNOPKVDPRDITSGNSSINTNTTTSRQVNDNHTPMSRHHVQDKPPTSI 2358
QY 3487 HBDLYTGKEISYNINNSTNTNDIPMANRDSYRGIDINDLSL 3530
Db 2359 HDRNLTYGEEYNYNVNM--VNTMODIPINSHNNYSSIDILNDTL 2401

RESULT 8
T28157
erythrocyte membrane protein 1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28157
 R:Yang, S.Q.; Wollish, W.S.; Gut, J.; Wu, J.; Ahn, J.; Petersen, C.; Fujioaka, H.; Aikawa
 submitted to the EMBL Data Library, July 1995
 A:Description: The molecular cloning and DNA sequence analysis of Plasmodium falciparum
 A:Reference number: 220479
 A:Accession: T28157
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2212 <YAN>
 A:Cross-References: EMBL:U31083; NID:g1517813; PID:g1517814; PIDN:AAB06961.1
 C:Genetics:

Query Match 13.8%; Score 2677; DB 2; Length 2212;
 Best Local Similarity 31.1%; Pred. No. 4.5e-117;
 Matches 713; Conservative 307; Mismatches 800; Indels 476; Gaps 83;

31 SHSARNVLERYAKNRHPSKAKEHVS-----LKGLDTAKFPGG-----PSRPVVK 79
 9 TNSAKEVLEDEIGETIQ-----KKAHSDADTFRSOLKNGEAKFYNGEIMOPNSKL-- 61
 80 HNYYPEPCNLDEKHTNLRKYDDVNLHPCHGHEONRFDEDESEC-GNKIRYKRRKDA 138
 62 -----CELDHTIDTNTVDGHSN---PCEGRQTVAFPDNRNSQCTKNRKH-DVDSNV 109
 139 IACAPRRRRHCKNLEALNDIN---TONHDLGVLTVAKEGESYVNNHP---HKGT 132
 110 GACAPRRRLHCKSHNLESIOITNNYDSSKAKHNLAECYAKFEGESIVYVNLQGHHTY 169
 193 SDACITLARSFADIGIVRGIDMEKFNVDH-----KYETGLREVFKKH----- 236
 170 EGCITLARSFADIGIIRKDIYLGNPQESARKOLENLRKIFDTIELTSSRNGKT 229
 237 DGMEDVKNYDNGSGNYKLEAMANNVNRNFWEAITCDASYKSGYFMOSESNTPLFS 296
 230 NGAEERYK-----DGSNGYKILREDMMNARLIDIMKAMICKAPGNAQYFNTCSNGEKP 284
 297 NPGCGHKGQVPTNLDYVQYLLWFDMEGEEFCRRIRIKAKYKDSR---NDKERLYCS 333
 285 GEGCOCIDGTPNPNDYVQYLLWFDMEGEEFCRRIRIKAKYKDSR---NDKERLYCS 344
 354 HNGHDTTITWKGILHDKRCDTSCKCYFVFWLNGOEAFFKOKREKYEKIOSYLSN 413
 345 RNDYDCTKTRISDIXSMRBECKTCLYDPIYKWDINKKEFEKOKKCKENELRYNNE 404
 414 DNKFVNNINSEYYKOFERYKLETOYATNDTFLNLNGEYKCGGLPDEKDTFTNSADK 473
 405 SQSPKRYNNMYETDFYGNLKKDYQSMND-FLKLINSEPTCTNIDKKRIDPTK--DPE 461
 474 GITYREYCOVCPDCGVK--CDGIKTYHKSDNDRERANNEDYRPMGVKFTNTITVLSGN 531
 462 EYFSHREYCDPCWCGKCKAQAD-TWKRLYENPQCPKIKRYEPPKGEETEDIVLTGK 520
 532 EGDITOKLENFCSNTYKDNKNKQWECYKDEINIRKLEONTLEINNPKIISPHNE 591
 521 ENNDIIVKLEBECKTGNIGFK-NEEMNCYQYGN-DKCYLNGESELGGRK-KYKDYDNF 577
 592 FELWVTVLLKDTIKWMDKLTCTGCTNNNTTTCIDECNRNCLCFDWMVQKKEEWSIKLFT 651
 578 LMFVVAHMLKDSIEMRSKSNCLSKDKCTCKKCNCKCYEKWIGKQVEMQIJKHFD 637
 652 KKKNIQ-QSYYSNINNLFEQ-YFFKWMKDKDKDEAKKELMENIKRKNFESNLNNDY 709
 638 KQDFQGWGMYFVLETVLEBDQFTTITKAYGAREIVHIOEMQKKREOV--LHEASN 695
 710 LENAIELLDHLEKATATC-----KDNNTNEACETSHN-----ATTYPCVPRGTOPTK 759
 696 MKTIIDELDHLEKAKKQCYVNHKDNKCPDSDSEDEEDIDPQRQPCAKPSSAARAL 755
 760 NIEIAQYFRSAVEEARNGL-HKLKGAHEGITYRGRGRKDPKDNLCRIMIKHSNRL 818
 756 -YAKVSNMHHKRRQLVNRGVSSKLGDAKGEYRKSRTTIKLD-ICISITDHSNAKR 813

819 GFSNGPCDCKGTGDIOTRFVVTGTEWVDEPHMKKHEDVIMPPRRRHITSNLEHQT 878
 814 GHDDQPKRRK---DSKSEMRTEDEMG-PAGFISKTYKDYMPRRORHCTSNLEHQT 869
 879 DHPANGTIV---DLVNSFLGDLVLSAKYEANKIIMYKKNLKKPKRVTPKHQTTI 935
 870 NKLINGNDIGNRRIINDSLGDLVLAANYEADFIKMYKKON-----DYKNATI 920
 936 CRAIRYSFADIGIIRGRDLMERNGDMVKLQGHLETVYKIHSLKNGDKYNDAPKY 995
 921 CRAMKYSFADIGIIRKDTMDWMDSGSKTRDLRELFIOTIKKKHPG1-KEYKEDRP-Y 978
 996 LKIRENMENAKWEMAMCDIKYLKOKSGHOSTOSYCYGSDHPDLDTYPOKLKMT 1055
 979 TKLRDEWEMANRRKRIWEMAC-----PTPNSSFPCKSHPTPLDITYPQRLKMT 1027
 1056 EMAEWCKYOKKEVDKLEKCKECKDKDKNGGCKESGCTCTEACNENYDIGLWKE 1115
 1028 EMAEWCKEOKYOGELVLSACNGCKDE--GRVCTNKS-SQCTSCMOCENYKNEINTWKE 1084
 1116 QWNIISDYKELHBOQOMSVNSGIEASSTAKNHDINVEFLSELYOQNGK----- 1168
 1085 QWDMKEIKYLVLYQAQTTAANGPDTYSGLVDENEKPPVNFLELYKENGKLGPNRT 1144
 1169 --SNKSGTSPESAVIGTNTYENVAYLH-DTGNFDQOSQNEFCDEKSPCKNEKAFR 1225
 1145 PRAKRSKRETPAPSAVANDYSTAAGYVHQBMGHMECKYQTERC-EKTEBOYENENTFK 1203
 1226 DKPOHDGACGCKSGSKPTVQJTKRK-----KAEKQTE----- 1260
 1204 NPPQYQDACICTRPPPKEDSKRSRSDSDEEKVETKYEKATEVADVTGPPAPKEA 1263
 1261 -----CKTVNDILKENDGKKOVEDCHPKKNSNGY-PDWOC-----GNT 1297
 1264 TTTLIDVCPYIAGVLT-----KENLENACPTKYGPAPRPSMKCIPTKTNATGSRGSSGNG 1319
 1298 NLYEDPR-----VCMPPRQKLY--HFLANDNEIKLOSQV----- 1332
 1320 ALDRAKATVSSGSPVTSNGSICIPRRRRRLYQKLMDASGTVVSGAQTPOGTS 1379
 1333 -----NLEKAFIKSAAEFFSWYKYSKDEGENELDKELKE----- 1369
 1380 PSKERTPSDLKRAFIOAIAIEFFLMDRYKKE--KELEKKEKAVANGVYPSLNGRP 1435
 1370 -----GKIPAFILSMFTYEGDYDPLFEGTD--ISKHGE----- 1402
 1436 PQCPGYTGDSPOSLAQGTGYIPEPLRQMEYTLGDAYADIFFGKNDIYIDRKNQKDAER 1495
 1403 GSKLEKIDSLFRNGQOKSNGKTRQEWATLHSHETIEAMALCY-----KI 1449
 1496 EKKIKDAIEVLNNAOSQPPSDEKROTWMQNGEHIWNGMICALTYYEKDEKGTPLKQNE 1555
 1450 GAK-----KDFETNYCYNNVYKFSKDS-----TTLFEFAKRPQPLWJTE 1489
 1556 GLASALMDENKKNPKKQYQYDKAKLDENSGTSPKTDHPPPLTNTFISRPYFRLEE 1615
 1490 WYDDYCYTRQKYLKDVQCKSSND--OLKCD-----TECNKKC 1525
 1616 WGEFFCEERKRRKLEKIEVECMDEDEGKKOKCSGDEDEEIRKODYSTVRDFYCEBCKYC 1675
 1526 EDYVYKMKKKK-EMIPQDKYKDEKDKKRRDRQHIGWYVDYGTNTADLNKFTASC- 1583
 1676 RPYKRWLEKKRDEYDKQEKYENNOKTARRNNNDNASTTLDVCTTGTAGDPLQTLKNGPK 1735
 1584 -----GDKPQASVYQNRIOLEKQAYYDADAKHCGCTFIENDK-----YTNISSKDKCKGL 1636
 1736 NDANVDSGEKKKIFDEGDFKTYQYGT---CSLNFCKNGDGCRRATRYVT-----CNGS 1788
 1637 VKEANTGAIWQNGKPNYNNLKELTEDVLPSPSRRLICFHALDGNATYDDEVYDENGKLR 1696
 1789 NRPTTITADDIKNGG--NSAEIINMLVSD-----DINGNGFN- 1823
 1697 RIMEVATIEGYNLGOYYKKEKKEKKEKITSDAHKSYEVPPCSAMKYSFYDLRLIL----- 1752

Db 1824 ---DLECKNANIKFKIKENMK-----CYFECKSDVGLKKNNDIDONQIILRAL 1872
 QY 1753 ---GIDNLEDEKOTEENLTKIFNKNGSVGKSGDSTTTNGSGARFFENNEKCYWNA 1809
 Db 1873 FKHLFELFDYNNKIRKKNLCINN-----GEKAITNGVEO-----WINQRTMT- 1920
 QY 1810 MTCYKRRGDDGNSGNSARSDDELKCGSVSPDDYPMGRNRDEGATAYOFLRMFAEMGED 1869
 Db 1921 -----NIKN--RFEEOY-----NGDDTEM--KSPFRSLVLDILROLA---AT 1955
 QY 1870 FCKHEKELEKIVA--CNDYTGDNEDKRRKCKDCTOYKFKFISEKPOYEKOIKKYE 1927
 Db 1956 IDKNHNGGLKIVKSVCN---CGNSONGKE-----GE 1986
 QY 1928 NKDKITSEHPVANDAEADAREYLDKOLKICEN--KSGDCEYKCMKDVSOTRLD----- 1979
 Db 1987 ENLDVYL-----LLOKLEKAKCKDNBPETSGIPQPC--EVSFPHIDEQPLE 2034
 QY 1980 --GNSQMPASLDD---EKEVEGKNCQVPPRPVRRPETSPPVSLISKATASKKEA 2033
 Db 2035 EEEVTEHPKICDDVLTQPOPEPEBETCBESGPTDVKKEEKEE-----KEEKDEE 2089
 QY 2034 KIAPPTKOPKVENLT 2049
 Db 2090 PGLPPTPPAPSPST 2105

RESULT 9
 T14602
 Variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
 C:Species: Plasmodium falciparum
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14602
 R:Voss, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
 Submitted to the EMBL Data Library, February 1998
 A:Description: Identification of a conserved 5' flanking region of Plasmodium falciparum
 A:Reference number: Z18158
 A:Accession: T14602
 A:Status: preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2135 <VOS>
 A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AAC05220.1
 A:Genetics: varph17
 A:Gene: varph17

Query Match 11.9%; Score 2303; DB 2; Length 2135;
 Best Local Similarity 29.4%; Pred. No. 1.4e-99;
 Matches 681; Conservative 306; Mismatches 745; Indels 584; Gaps 95;

21 GDAKSPITKESHKSARNVLERYAKNI--RHSRYAKEHYDSLKGLDTLKAFFGGSPSTPVNK 79
 Db 5 GPOGGGTRKD--DAKHMFDRIGTIGKQVHNAAEFEFNELOGRSSA-----TNTNL 55
 QY 80 HNYYYRYCNDLHK--ERTNLRTDVTNLRHPOHGREONRFDEESE--CGNKIRNYKRN 136
 Db 56 KLIGNIETCSLEKYYKYPNGGVSQDKRYP--KELSRKYVDPRSDKIGGQCTNEKMR 113
 QY 137 DAL--ACAPRRRHMDKMLAL--NDINTONI--HDLGNVLTAKYBESE--VNNPH 189
 Db 114 DGIAGCAVRLHLCHNLLETETNNYSNNAKHLVDVCMARKYEDSITKYTTGHOH 173
 QY 190 K-GTSDACTALARSFADIGDVRGIDMFKPVNHDK-----VETGLREYFKIHDGMEV 243
 Db 174 KYDSOLCTVLARSFADIGDVRGIDLFYGTQOKREKREDLENLKDLFKIHSGLIGCV 233
 QY 244 KNDNPOSSGGYYKLRKMMVANKYWEATLTCASTKSGYEM-----QSESNTPLFSNP 298
 Db 234 KARYK--DTNFEYELRDMWMTANREYWKALTTCDA--HGTYFRATCGGDNESPMAKNKC 289
 QY 299 KCGHROGK---VPTNLDYVPOYLRFMFDEMGEEFCRKRNIKLKVKVQSCR--NDKRELY 351
 Db 290 RCGKKDGRPDQVPTFYDYVQFLRMFEENAEDECRILKRLKLDATQCKREKIKGAKKLY 349

QY 352 CSNHGDCSTTTIKKILHLNKCCTDSTKCYFEVWLGNOQEAFAKQEKYEKETOYL 411
 Db 350 CDLNRYDEQATSGKHDFFEEDECKGCOYSCARFVWIDNOKREFLKQKKKTGTEL-SLK 408
 QY 412 SMDNRVNNINSXY-----KQYEKLEFOYATNTFLMLDEGKYC--KGLPGEKDI 464
 Db 409 SRKRKAGGISTIKVYGYEKFEKLEK--SEYRVGEFLGLNNETCKEVEKDKKGGKIDF 467
 QY 465 TETNSADKQ-----TFRSEYCOVPCDGCJIKYTHKSDNDRERNVNDPRPMGV 519
 Db 468 KTVNSGASAGDVNKTFFYTKYCEACPWGAEQERNGVMAKADRDGSPGDYK---Y 524
 QY 520 KPTNIVYLSNGEGDITQKLENFC-----NSSTNYKD 552
 Db 525 KKEEITLIGDYTKSEIYERYKFKCKNKGANGREGVGSGEANGAASNSDNATTYGCG 584
 QY 553 KNN-----QKWECCY--KDN-----INRCKLEQNTETINDNPKTISPHNEELMYTL 599
 Db 585 GGNDSLSLEKWTQYKKKKENNDGKDINFC--VQAMONSKKQKQVKSINAFMDWHDH 643
 QY 600 LADITKMDKLKTCIN--NTTHCIDECNRNCLCPDRWYKOEEREMSIKILFTKKNIQ 657
 Db 644 LIDSIKWRNEHGKCIKNKGKT--CIKCKSKDCFLKVVQKEKEMKILHEFTNGGFD 702
 QY 658 QSYNSINNNLFEGYEFKYMDDKDEAKWKELMEIKKKKESLNNRDYLE----- 711
 Db 703 KGHORLGFTHDVYLYNLIDK-----KELKTI--EGTYGNTERTKIKKELDKEE 751
 QY 712 -----NALELLDLHLETAATICKDN-----TNEACESH 741
 Db 752 KDAGTGVAAGTGPRKNIMDKLIEHLEQAKCKQOEPOSLGSLNHHVYDDGSPKR 811
 QY 742 NATNPCKYKPRGQOPTNIKEIAQYF-----KRSAYEARNRGLHK--LKGAEHGIYKR 795
 Db 812 DRTNPGCSDT--TEYAVLAGVAKQKQGEVARKMLERSRKNKTSLESDIKKAQFN 870
 QY 796 GGRKDFDNLRLIMIKHSNRLG--FSNGPCDGGGTGIGTIRVYGTGEMVDEPHKRD 854
 Db 871 GRSGELNGDICKIDNKYSNDIRSTAGPCTGK--DQGERRNAQTKWBGD--NFSAT 926
 QY 855 HEDVIMPPRRHHTSINLHLOTDDHPLNGNIYDVLVNSFLGDVLLSAYEANKIIRMY 914
 Db 927 HKNLYIPRRQHMCTSNLEKLDLVSYSKN-----VDSFLGDVLLAANNEAQRTDHF 981
 QY 915 KKNLNLGPKKEVTPDKKHQTTICRAIRYSPADIGDITIRGRLMENGSMVLOGLLEVFG 974
 Db 982 AHKKDDHG-----IACRSVRSFADLADILIRGRMDKDDGAQKMEIDFKFIIG 1030
 QY 975 NIHSKLG--KNDKYND--APRYLRLRENWMEANRAKYWEAMKCDIKYLKDKSGHST 1030
 Db 1031 NLVSLFLGKING--KYDGDQDRTPOYKQLRBDWMEANDQYWKAMVC-----EKDGKCD 1082
 QY 1031 QSSYCGYSDHPLDDYIPOLIRMTWTEAWYCYQKQEKYKLEKCECKDKNGGCTK 1090
 Db 1083 E-----DPTPVDDYIPOLIRMTWTEAWYCYQVQSEYDELKCKGSCCKIKKQVGC-- 1134
 QY 1091 ESGTGCTKCTEACENYNDIGLWKEQNNISDYKLEHQAQSVNSGSEASTAKNHI 1150
 Db 1135 SGSDSDCTPCAELCTTYGQKIKPWEDQNNMLLOTYLLWQAEFTTAYGGRAYSGVCGK 1194
 QY 1151 DRNVIELSLYOQNGSKSKSGTSDESAYVGTNTYENYAVLHDTGNFDDQSOQNEFC 1210
 Db 1195 DKPVQGLELELOKONSQK-----TYNNPAAAGYIHQEARVGECEVQKYFC 1238
 QY 1211 DEKSDGKDNKCYAFRKPPOHDGAGCGK-----SGSK-----PTVQITKKAEE 1256
 Db 1239 NTNG-----NODKYFRKPKDHDACCKADPOQASAGGAGARLSPRPVSDDDHSSD 1295
 QY 1257 KQTE----- 1265
 Db 1296 EDEEEDDEDDGDAEDENDEPASEEYKDDTEDVVEETAVSOPAAPTTTTPGVTPACEIYK 1355

QY 829 GTGDIOTRFVGTWEVDPHEMRKDHEDVIMPRRRHICTSNLEHLOTDPHNGNIYD 888
 Db 764 GYGD---VKELKJIKELDEEKKKKQAEAVV-----VV- 794
 QY 889 DLVNNSLGVLISAKYBANKIIMYKEKNNLKPKREYTDPKHOTTCIRAIRSPADIGD 948
 Db 795 -----VAONOKKTTIDKLQHE----- 811
 QY 949 IIRGRDIMERNGDMVKLOGLHLETFVGNHSLKGNKNDKYNDDAPKYLKRENNWEANRA 1008
 Db 812 -----GDDANNLKTH----- 822
 QY 1009 KWEAMKCDIKYLDKSGHOSTOSSYCGYSDHPDLDYIPQKLMTEMAEWYCKYQKE 1068
 Db 823 ----- 822
 QY 1069 YDKLEKCEKCKDKD---NOCGTRESGTCTCTEACNEYNIDIGMKRQNNISDKY 1125
 Db 823 ---KECCEFTOPKPPAGPGAPSETGETT-----LEDEE 856
 1126 ELHQAQMSVNSGIEASSTAKNHIDRNVIEFLSELYOQNGKSNKSGTDESAVIGTNT 1185
 Db 857 EEDBE---EDAGDEV-----EBGETVDT-- 876
 QY 1186 TYENVGAYLHDGNFDDCOSQNEFCDEKSDGKNEKTAFRDKPODHDGACGCKSGSKPTR 1245
 Db 877 -----TEGDETFVEOPVKOTDREG----- 896
 QY 1246 VOJTKTKKAEEKDTE---CTVNDILKENDGKQOVEDCHPKKNSNG---YPMQC---GN 1296
 Db 897 -EBEEKAKKADTTSLDVCCTVKNALTNND--NLTDACLKATGPGSKKEFPNMKVCYSSE 953
 QY 1297 INLV-----EDRVCCPPRRKOLCYHFLANDNEIKKLOSQVNLKAPFISAAA 1344
 Db 954 KSVATGSSGAGCKSGDKAICVPPRRRLYYGGLK-----LTSACT 996
 QY 1345 EEFWYMYKSKDGEENELDKELKEGKIPPAFLRSMFYTFGDPYRDLFTGDISGGBES 1404
 Db 997 SS-----ESPQG--GSESSR-----ASDSQONG-- 1018
 QY 1405 KLEQOISLFFKNGDQKSPNGKTRQEWTEHSHIWEAMCALVKIGAKKDPTEYNGYNN 1464
 Db 1019 ----- 1018
 QY 1465 VFSKSTLEFEFAKRPQFLRWLTEMVYDQYTRQKYLKQVEKCKSNDOLKCTECKNK 1524
 Db 1019 ---GDDITTESLRK-----WFI-----TAA 1037
 1525 CEDYKMYKKKKEWIPQDKYKYDERDKRFRQIHGVVYDTGTNATDYLNRKFTASC 1584
 Db 1038 IETFLMHRYKKEMWAOKK----- 1056
 QY 1585 DKPGASVVOHRIOLLEQAYVYDADKHGCGIKFIENDKYNISCKDKGLVEANTGA 1644
 Db 1057 ---AELORNGLL-----GTA 1070
 QY 1645 IKWQKGNNTNNLKELTEDVLFPSRLRICEHALDGYTDPREYK---DENGKRLRME 1700
 Db 1071 S--JMLGGDSDSPQTOLOKSGTIPLDRLMFTYU--GORYOILVGVADDKNG-- 1120
 QY 1701 VAAETGYVLGQYKKEKKEKIKISDAHKIYSEVPPCSAMKYSFYDLNDITLIDNLEDE 1760
 Db 1121 ---GNN-----ITLMSGNKDE 1134
 QY 1761 KQTE---ENLKTIFFNKNGTSVSGSDSTGNGPSTARKFFMNNENKCYNNAMICG--YK 1815
 Db 1135 KQMEKIEKIEQILFTPSGNKETRGPONSVD---ROSLMDRIAHEHVMHGMCAJLYK 1189
 QY 1816 RGRDGSNGSARSDEDLKCKGVSDDDYPMGNRDEGTAYOPLRWFPAEWGDFCKHKE 1875
 Db 1190 ---DDDNQ----- 1194
 QY 1876 KLEKIVGACNDYTCGNEDKRRKCTACTQYKFISEMFPQYETKQIKKYGEMKDIYSE 1935

Db 1195 ----- 1194
 QY 1936 HPVAKDAEDAREYLDQOLKICENKSGDCEYKQOMQVSNQRLTDGNSQMMASLDERKE 1995
 Db 1195 -----LKV----- 1198
 QY 1996 VEECKCQYPRGPVRVRETPSPRVSLSKATASKREAKTAPTRQPKVENITTEBRAQ 2055
 Db 1199 -----YKPKXIKEN----- 1207
 QY 2056 TTRBRAAQOTRKRTSTATTTESDVGTMVAKILSNKPDSCRGIEGCPKTYGOYPRWGIV 2115
 Db 1208 ----- 1207
 QY 2116 GSKENENGICMPPRKICLNNIOLYNETENKRDNDIKAEAFICAALETQFLMKYII 2175
 Db 1208 ---PEK-----LM-- 1212
 QY 2176 ENPAENELQNGTIDFERKRMYYTYGYKDMFEGTISNDKIIITVNSVYTLNENNK 2235
 Db 1213 -----NETTK 1217
 QY 2236 KRODKDEBELKIFWEKNNKFIWEGMUYGLYHILTDENEKERTDNYQYDMTKLPSL 2295
 Db 1218 KPKDEKYQYQAKL-----EDESSEKRPDSASG--TKLT-- 1250
 QY 2296 BEFYRPOFLMFTWMAEFCKNRKREOLKLEAGCKEYECGNSDGTQOECAG-- 2348
 Db 1251 -DFIKRPYFRLYEBWGENEFCFKRTMELG IKEDC--YKNGRSGSGGLKONEVYIDKEK 1307
 QY 2349 ---ACLYONFIKKMTEYERQREKFKKDKGKYDYDSTERDILEKATCAHEYL 2400
 Db 1308 IFGDLICTCARHCFYKMW-----INTKRD----- 1333
 QY 2401 NMLKELCGNDCGOMQPSQOLPKTQOSQSDANDMPESLDVYPEFKNCEPBLSKR 2460
 Db 1334 -----EFNK----- 1337
 QY 2461 GSMHTKKITEPKIPMNCVEKAAAYLSKEANNMDITLKEFPIESTKESKESNWTNN 2520
 Db 1338 ---OSNAYSQOKKY-----EENDSAQKN 1360
 QY 2521 NPCDKKPYAPDKYIGR--NPPCENRENEFKYDYEMKCYKNSKFOEKKRVVPPRRRH 2578
 Db 1361 GVCGLTDDAAE--FLNLKNGPCNKSEBENKKADELIDFKKPDPTFRADN--CKP-- 1413
 QY 2579 MCLBNLDEIKIERLKDSNYLLKAVRTARNEGJDIIFNFSENGAMPICD--TKRYSFA 2637
 Db 1414 ---GSEFKIK-----CENHNCSSGNGTQCKCDKTTIANT 1445
 QY 2638 DLGDIYRGST---DMLRIGVYLPVEYIKLYKVFYIYIGKMNKKNKKNKNDVOTFPSANW 2694
 Db 1446 EIEINKTNTKEVIMLVSDDSKATEK-----DGLSECKDKGIFGI-- 1487
 QY 2695 DANBKDIWAKMTCKAPBDAKLFRKGRMDGFERITL--IODKCGHKDPPVDDYIPORFRW 2753
 Db 1488 ---RKDEWEC-----GKVCYGDICNLKKNKONIGESD--KXITIMK--ELL 1526
 QY 2754 TEMSEYCYKAMELEKFKKSDCHCTSDRCANDYDENKCEOCKTRCOEYKNEFVLKWSL 2813
 Db 1527 KRWLETF---LEDYKNIKIKHSHCTKNGK-----GSKC-----IKGCVKMW-- 1564
 QY 2814 FDIQSNKYELVBOPIYTRISTYDHAQNVOKLTKPKSCSVSESEYELHESKLANKF 2873
 Db 1565 -----VQOKKEMKQK-----ERNE-----OYKS 1585
 QY 2874 NENDGSSNIRITAFETTPKSYKACSCYLPKSNPLDNCPTDQNKDGKELQTFPCSRND 2933
 Db 1586 KTSDEYFNKSFLETWIPK-----IAVYND 1610
 QY 2934 YDNNDLNMNAAVYVLSNDDNKGVLIPRRRHICTRPIYANVRKGDKELLKLLTSAFS 2993

Db 1611 QDN-----VILSKFGNSCG-----CSASALSTN-----GNEE-----1638
QY 2994 QGOLLQOKKSEELCFEAMKITYADYSDIIKCTMDMDISLSEKIKIIFTSSEATEENRK 3053
Db 1639 -----DAIDCMIKLEKKIDCEKCKRPGEN--1662
QY 3054 TWMENNRQIWMHMLGCKYKATSKVTLDEGWCQLPKDEETNOFLRMLEIAKQACKKKH 3113
Db 1663 -----SGQTC-----1667
QY 3114 VDSLKTCPRSNEDNFEASELLROPGCCONDIRKYLISLILIKTMENTENLIKQYKQLKDS 3173
Db 1668 -----NELTLP-----1674
QY 3174 SGNDKPEEENVOYIKSKDSQALENDINEIVTGTKNNENKEVELKLYPGLY-F 3232
Db 1675 -----LDVODEDEPLEETEENP-----VGKHPSCEP 1702
QY 3233 VEDETHKNHVLDOGNKEE-QIVRPKALYFPFHVSFYAPLFSYHRAQYDPKND-I 3289
Db 1703 VEDKK-----KEEGEETCTPAS-----PAPAPAPASPSTPAPADEPEDPTI 1746
QY 3290 LKSSIVVIYALGLIALHEMK-----KKFKSSVDLRLINLPQG 3329
Db 1747 LQTTIFLGIALAGIAFLFLKAYIYICVYMYTMCPCITMKTKRHPVULFVYINLPKS 1806
QY 3330 EYGMPLSKNRYIPYRSGPYKGYIYMEGDT--SGDEKXMYMDLSSSDIT-SSESEY 3385
Db 1807 DYDIPKLSFNRYIPYTSKGRKRYIYLEGDSGTDSGYDHY-----SDITSSSESEY 1860
QY 3386 EELDINDIYVGSPPKTYLLEVLEPSSKRDIPSDPDS--NMPRTNREIDEMWNLKH 3442
Db 1861 EEMDINDIYVGSPPKTYLLEVLEPSSKRDIPSDPDS--NMPRTNREIDEMWNLKH 1920
QY 3443 DEVSQYLPNTEPN--NNYSADIPNTEPNTLYSDNPEKPTIISHDRDLYGKEISYN 3500
Db 1921 DFTSMNLQNPQVNDYKSGDIPFNTQPTLYEDKPEEKPFITSIHDRMLNGEESYN 1980
QY 3501 INNSTNTNDIPMARNDSRGIDLINDSL 3530
Db 1981 VNMSTNSMD-PKYSNNVYSGLINDSL 2009

RESULT 11
T28634
variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
Accession: T28634
Su, X.Z.; Heatkole, V.M.; Wertheimer, S.P.; Gulnet, F.; Herfelde, J.A.; Peterson, D.S.
Cell 62, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
A:Reference number: Z20487; M0ID:95330813; PMID:7606788
A:Accession: T28634
A:Status: preliminary; translated from GB/EMBL/DDJ
A:Molecule type: DNA
A:Residues: 1-2182 <S>
A:Cross-References: EMBL:L42636; NID:9886379; PID:9886380; PIDN:AA75399.1
C:Genetics:
A:Note: var-7

Query Match 9 0%; Score 1743; DB 2; Length 2182;
Best Local Similarity 20.3%; Pred. No. 2.2e-73;
Matches 757; Conservative 339; Mismatches 678; Indels 1960; Gaps 126;

QY 13 GNAASLEGDAKSPITKESHK-ARNVLER-----YAKNIRHPSKYAKEHVDLKGDLTK 66
Db 10 GSGGSS-SGCGKMDTSYIYVSDAKDLDDVGEKYEAKKND-AKKTIEALKGMINT 66
QY 67 AEFRGGSSTPVNKHNYYPYPCNLDHKEHTNLRYDVN--LRHPC--GREONRFE 119
Db 67 ANGRSSTASSIE-----TCTL-VKEY-----YERVNGDGKRHPCKDAKNEDVNRFS 114

QY 120 DEESEC-NKIRNYKKNDAI-ACAPRRRHMDCKNLEALNDINTONIHDLGNVLYTAK 177
Db 115 TLGGQCTYNNRIKDSQGDNDVGCAPYRHLHDYNNESID--TSTPHLLLEVCAAK 172
QY 178 YEGESIYNNH-PHKT-----SDACTALARSFADIGIYRGIDM-----KPNVADKY 224
Db 173 YEGNSI-NTHYHQHQRNEDSASQLCTVARSFADIGIYRGADLYGYDNKEKEQKKL 231
QY 225 ETLREYFKKIHGDMEDEVAND-----YNDPG-SGNYTKLEAMWNNRKKWEAITCDA 278
Db 232 EOKLKIDIFKKIH--KDMKTNGAQERYIDAKGDFEQLREDMWTNSREYVWALICHA 288
QY 279 SYSGYFMQSESNTPLESNKCKGKQKVPNLDYVQYLRWDEGECECRKNILYK 338
Db 289 PKANFYKIKTACVNGKCTNOCHCIGDVPYTPYQYLRWEEEMADECRKKALEN 348
QY 339 VKDSRNDKRLVCSHGHQCTTTMKKGLIHLDNKTCDSTCKCYEVMGNOQAEFK 398
Db 349 LOKOCRDYEQNLVCSGNGYCTKTYIKKGLVIGEHCTNCSVWCMTETMIDNOKKEFLK 408
QY 399 OKERYKEI-----OSYLSNDKRFVNNINSEYKQFYKLEKTOYATN 441
Db 409 OKRYETEISGGSGSKSPKRTKRAARSSSSDDN-----GYESFKYKKLEKVEGYQDV 460
QY 442 DTFNLNLNEKCYCKG---GLPGEKIDTFINSADKGIPTREYQYQVPCQKACDGIKY 497
Db 461 DKFLKILNKEGICOKOPQOVENKADNVDTNEKYK--TFSITLCEPCPCMGLEKGGPPW 519
QY 498 THKSDNRERNVNDYKPRPGVKTNTVLYSGNEGODIQKLEFNCNSSTNYDKNNOK 557
Db 520 KYGDKTCGSAKRTYEPK---NTTIDLPVLVLPKSOQNLILKRYNFEKCAP-GGGQIK 575
QY 558 WEQYKD---ENINCKLEQNTIINDNPKIISFNFEWLYTYLLRDTIKMNDLKT 612
Db 576 WOCYDYHRSKNNKNCVGTWMDKFTQKQYKSYNVFEDVHDLHDSVEKTELK 635
QY 613 CINNTT---HCDEBNRNCIGDRAVYKOKKEEMNSIKLFTKKNHIOQSYSNINLP 668
Db 636 CINNTNGNTCRNNKCKTDCGCFQKWEKKQDQWMAIKHFGKQTDIVQ--KGLI 690
QY 669 EGYEFKMDLKDDEAKWEKLEMINIKRKNESMLNNRDLT-----NAI 714
Db 691 VFSYGYLDLV---LKGNLQNIKDYHGDTPDIKIKILLDEDAVAVLGSKDWTI 746
QY 715 ELLDLKETATTCCKDNTNEACETSHNATNTPCVKRGGTOPTKIKEIAQYKRSAYE 774
Db 747 DKLLQHEKEQAEQCKR--QEECE-----KRAQ 773
QY 775 EARRGLAKLKGAHEGTYRGRKRPDNLCHIMIKHSNRNLGFSNGPCDGGTGDGI 834
Db 774 ESRRRSAAE-----TREDER-----TOOPADSAGE--797
QY 835 QTRFVGTWEVDEPHARKDEHDIYMPRRRHICTSLNHLQTDHPLNGIYVDLVNNS 894
Db 798 -----VEEEDDDDDVDEDED-----DDVQEE 820
QY 895 FLGDVLLSAAYEANKIIRMTKEKNNLGPREVTDPKHQTTCIRAIRYSFADIGIINGRD 954
Db 821 EEG-----KEEGTYEVEVEVEET-----842
QY 955 LMEWNGDMVQLGHLTFVPGNIHKSILKGGKNDKYNDAAPYKILREMMWENAKKVEAM 1014
Db 843 -----842
QY 1015 KCDIKYLLKDSHQSTQSSYGYSDHTPLDDYIPQKLMMTEMAEWYCKVQKREYDLKE 1074
Db 843 -----VTE-----OE 847
QY 1075 KCREKDKDNGQCGTKESSGTCTCTEACNEYNDIIGLMEQWNIISDKYTELHEQAOMS 1134
Db 848 GVKPC-----DIVKCL-----858
QY 1135 VNSGICIASSTAKNHIIDRNVIEFLSELTYQONGSKNSGTSDESAVIGTMTTYENCAYL 1194

Db 859 ----- 858
QY 1195 HDTGNFDDCOSONEFCDKSDGKDNEMKAFRDKPODHDGAGCGSGSKPTIRVOJTKTKKA 1234
Db 859 ----- 875
QY 1255 EEKTECKTVNDILKENDGKQVBDCHKNNNGYPRMOCGNILV-----EDPRVCM 1308
Db 876 ----- 906
QY 1309 PROKLCVHFLA-----NDNEIKLOSov-----NIKEFIKSAAEFFESWY 1353
Db 907 PRRRRLVYGGISOMASNGGDETTESAVSAPSQSESEKRLTAFAISAAIEFFELMHKY 966
QY 1354 K-----SKDG-----EGNELDKELKE-GKIPPAFLRSMEYTFEGYRDLFLG 1393
Db 967 KEERKPPATODGAGLGVSLPEPPSPGEDPQTOLOQTGVIPDPDLRQMFYTLADYKIDL- 1024
QY 1394 TDISKGHGBSKLKEQIDSLFKNGDOKSPNGKTRQEWTEHSEHWEAMLCALVKIGAKK 1453
Db 1025 ----- 1024
QY 1454 DDEFENYNNVAFSDKSTLLEPAKPOFLRWLTEMWIDYCTTOKYLKDVOEKCKSND 1513
Db 1025 ----- 1024
QY 1514 QLKCDTECNKCCEDYVKKMKKKKREWIPODKYKDERDKRFRQHIGVMYDTGTNAD 1573
Db 1025 ----- 1032
QY 1574 YLNRKFTASCGDKPGSASVYQNRNIOLEQAYYDADKHGCGTKFTINDKYNISKDKC 1633
Db 1033 ----- 1047
QY 1634 KGLVKAANTGAIKMONKGNNNNLKELTEDVLFPSRLRIGFHALDGYTDPVYDENG 1693
Db 1048 KNLVLEAS----- 1055
QY 1694 LRRLMEVATEGYNLGOYYKKEKKEKIKTSDAHKYSIEVPCSAMKSYFDLIDLIG 1753
Db 1056 ----- 1066
QY 1754 IDNLEDEKQTEENLKIFPNKNGTSYKGSSTGTGNGSTARKFPNNENKBEVNNAMIG 1813
Db 1067 ----- 1113
QY 1814 --YKRGRODNGNSNARSDEDLKCGSVPSDDYPMGNRDEGTAYOFLRWFPAWGEDFC 1871
Db 1114 LTYKE-----NDAR-----GTS----- 1125
QY 1872 KHKEKELEKIVGACNDYCGNEDKRRKCTDCTOYKFFISEMKQYKQIKKYGKDK 1931
Db 1126 ----- 1133
QY 1932 IYSEHPVAKADAEAREYLDKOLKIKCENKSGDCFKCKDVSTORLTGNSONMPASLDD 1991
Db 1134 ----- 1147
QY 1992 EPKEVEGKNCQVPRGPRVRREPRSPVSLISKATASKKAKTAPPTKQPKVENILTE 2051
Db 1148 ----- 1147
QY 2052 MRAOTRRRAAOQTRKRTSTATTESDVTWVKAILSNKPSRSGISGCNKTGYOPKW 2111
Db 1148 ----- 1147
QY 2112 GCIVGSKENENGICMPRRKKLCLINNIQYINTEENKRDNDIKAFIKCAALFTOFLWL 2171
Db 1148 ----- 1159
QY 2172 KYIENPAENELONGTIPDEFKRIMTYTGYDYKDMFEGTIDSNKKIITVTSVTIILN 2231
Db 1160 ----- 1160
QY 2232 ENNRKQDKKDEBELKIFEMKNNKFIWEGMIYGLYHILTDENEKEKINDYNDMYL 2291
Db 1161 ESGAKSND----- 1171
QY 2292 TPLSEFVYKPOFLRMTWAEFCNKRREOLKLEAGC-----KEYCNGSNDK--- 2342
Db 1172 PPLTKNFVEIPTEFFRWLHMGNSFCFERAKRLAQLKHECMDEBGKQY---SGGEJCE 1227
QY 2343 -----TOCAEACTYONFETIKKTELEERQREKREKQDKKYIDYSTERD 2389
Db 1228 ELPFSKQYVNILODLSSCAPCLYKWLIEKKTELEKQOKAYEQDQ---SNVENEKD 1282
QY 2390 IEKATCAHEYLNMKLKEILCNGNDCSOMQPSQLEPRTTQOSQSDANDMPESLDVPR--E 2447
Db 1283 -----KC-----QTSNNANANFSTLGAASP1A 1306
QY 2448 EFNKCCPELSSKSGSMIHKKITTEPKIPANCYERAKAYLSKAEANNMDITLKEFIPES 2507
Db 1307 EF-----LOKLGS-----CKNDGY-----ENGEDNKIDFK-NPDKT 1337
QY 2508 TREKESKSWTNNNPCDPKPYAPDPKYIGRRNPCENRENFKYDYEMKCKYKNSKFTQEK 2567
Db 1338 FKEAHS-----CDP-CRITGVK-----CONGH----- 1358
QY 2568 KRVCPYPRREHMCNLNDELKIERLKDSONYLLKAVRRTAREGDIJIKNFSENGCAMP 2627
Db 1359 ---CVSANGRECKNN--KITAEDIKN---KIDPAGNILEMYVSDSTN--- 1398
QY 2628 ICDJMKYFADIGDVRGTDMLRIGYVLPVEIKLYVFEYIYKKNKKNKGRNKYDVQ 2687
Db 1399 -----TFENLGD-----CASSGIFKGI----- 1415
QY 2688 TFRSAMWDAKNDIKW-AMTKAPADAKLFRKGMDFERITLIOD-KCGHKDDPVDY 2745
Db 1416 -----KDEMKCANVC-----CVDICTLEKKIKNOGED--KRY 1447
QY 2746 IPORFRMTWSEYYCYCALMELEKFKKSDHC-KTSDRKNYDEKCKQCKTROQERY 2804
Db 1448 ITMK-ELLKRLLEF---LEDYNRIRKKIKLCTKREDDGK---CJRGJIE-- 1490
QY 2805 NEVLKMSLFDIOGNKYKLEYEOPITYKISTYDHOVQKLTQFKSECSVESFSEIHE 2864
Db 1491 ---KW-----VQ---EKTEMO----- 1501
QY 2865 TSKLANKENNOSSNIRTYAPETPKYKACSGCTLPSKNPLDNCPTDONKDGKEIQ 2924
Db 1502 -----KIND-----TY-----LE 1509
QY 2925 TFFPCKNDYDNNLMDNNAYLVLNSSDOKGVLIIPRRRLHCTRPITAYNRKDGKEIIL 2984
Db 1510 QY---KNDGNTLTN-----LEOFQYR----- 1529
QY 2985 KLLTSAFSOGOLIGOKYKSEEBELCEAMKYSYADYSDIIGKTDMDJTSIEKIKKIFET 3044
Db 1530 -----TERKNMIPKPDGID----- 1543
QY 3045 SNEATEWRKTWENNRRQIWMHMLCGYKATATSVYTLIDEGMQLPKDEETNOFLRLIEMA 3104
Db 1544 ----- 1545
QY 3105 KOACKEKHVSDLSKTKPCRSNEDNEFASSELLROPCCONDIRKYSILNLIKMENTENI 3164
Db 1546 -----KTSQGLNSTDN-----SQNSNNND-----LVCLLKLQK 1575
QY 3165 KYKOLDOSSGNIDNKPSEBNVQSYIKSDOSCALFLNDINELVGTKNENN----- 3217
Db 1576 KISECKEQJSGOTQ-PCQ---NSSLSGKSTLVEDVDYEE---ONPEKRYEQPKFC 1626
QY 3218 -EKEVLYKLYLGLYVEDETHKNHVLIDGNI--KEEEOQVRKALYFTPHVDSYQAPL 3274
Db 1627 PDWKBPCKENDEBVGICGDEKCKKVEYDSVITQKEEBAASAPESPLTP-----EAPK 1680

Db 970 TSGSTPPDSEKELKAFVESAAIEFFLWHRKYEKKAQAAGAGGLPVEGSEPREYD 1029
Qy 1364 -DREKKGKIPPAFLRSEMEFTPDYRDLFTGDISKGGHGSALKEDIDSLFKNGDQKSP 1422
Db 1030 PERKIRGKIPDGLRDMFTLGDYRDLFT----- 1059
Qy 1423 NGKTRDMWTEHSHEIWEAMICALVKGAKKDDFTENYGNVYKPSDKSTLLEFAKRPQ 1482
Db 1060 ----- 1059
Qy 1483 FLRWLIEWYDYCYTRKYLADYQEKCKSNDOLKCDTECNKCEDYVYTKKKKKEMIPQD 1542
Db 1060 ----- 1059
Qy 1543 KYKDERDKRRFDROHIGVAVTDYTGNTADYLNKRFKASGCGKPGSASVYQONIQLEK 1602
Db 1060 ----- 1077
Qy 1603 QAVYDADKHOCCKPFIENDDKYTNISSKCKGLVEANTGAIKWONKGPNNYNLIKELT 1662
Db 1078 ----- 1084
Qy 1663 EDVLFPSRLRICFHALDGNVTDPEYKDENGLRKRLMEVAATEGYNLQYKEKEKEKI 1722
Db 1085 ----- 1098
Qy 1723 KTSDAHKIYEVPCSAKMYSFYDLIDILGIDNLEDEKQTEENLKLFNKGTSGVGK 1782
Db 1099 ----- 1115
Qy 1783 SUSTGNPGSTARKEPWNKECVNNAMICGYKRGDRDGNCSNARSDEDLKCKGSPSD 1842
Db 1116 APLVLSHP-----QIWMENNGKTIWGHVYAL----- 1142
Qy 1843 DDYPMGKNDEGTAYOFLMFAEMGEDFCRKEKELEKLVGACNDYTGCDNEDKRRKCTD 1902
Db 1143 ----- 1142
Qy 1903 ACTQYKFISEKPOYEKOIKKYGENKDIYSEHVAKDAEDAREYLDKOLKICENKSG 1962
Db 1143 ----- 1153
Qy 1963 DCEYKCMKVSTORLTDNSONMPASLDPEYEGKNCQVPRGPRVYRRETPSPRSL 2022
Db 1154 ----- 1153
Qy 2023 ISKATSKKEAKTAPPTQOPKAVENLTTEMRQOTRRRAOQTRKRISTATTTESDVGT 2082
Db 1154 ----- 1161
Qy 2083 VKAILSNRPDSRGIEGCPNPTGYQYPRKMGCIYVGSKENENGICMPRRKLCINNIOYL 2142
Db 1162 ----- 1181
Qy 2143 NYETENKRDNDIKAFIKCAIETOFMLKXYIENPAENELQNGTTPDEKRIIMYYTGG 2202
Db 1182 NVKLD----- 1197
Qy 2203 DYKMFEGTDSNDKIIITVNSVTIINENKCKKODKDELRKIFWEKNKFIWEGM 2262
Db 1198 ----- 1208
Qy 2263 IYGLTJHLTDENEKEKIRDNQYNDMTKLPSLEFYKRPQFLWTEMAEPCNKKKEQ 2322
Db 1209 ----- 1236
Qy 2323 LKKEACCKEYECNG-----SNDGTOE----- 2360
Db 1237 LKQIKYVCKYVE--NGDYGRSGSDGACDSISTHDYSTVPFNCPCGCKHCSSTRAKWIERR 1294
Qy 2361 KTEVEROREKPKKODKXKXKDPSTERDLEKATCAHEIYNNMLKELCGKNDSCMQKPS 2420

Db 1295 KIEFKQSNAYQO-----OKTATRNGNTFDEKFC----- 1325
Qy 2421 SOLPRTQOSSDANDPESIDYVPEENKCECELSSKSGMHTKTKTEPKIPMNCVE 2480
Db 1326 ----- 1333
Qy 2481 KAAVYLSKAEANNMOTLKKEFIPLESTKEKSKSWTNNPCDPRKYPADKYIGRRNP 2540
Db 1334 -AKFLER----- 1355
Qy 2541 CENRENEFKVDYEWKCYKNSKFEQEKRCVCP--PREHMLRNLDEKIERLDSNTL 2599
Db 1356 ----- 1362
Qy 2600 KAVRTARNEGIDITIKFNSENG-----CAMPICD----- 2644
Db 1383 ----- 1420
Qy 2645 GTDMLRIGGLPYVEIKLYKVEYIYGKWRNKNNGRNYNDVOTE-----RSAMWDAN-- 2697
Db 1421 VSD----- 1444
Qy 2698 ---RKDIKAMTCAKABEDAKLFRGRMDGE--RITLIQDCKGHKDDPYVDYIPORF-- 2750
Db 1445 KGRKDYVKC-----GYCGVDICBQTNINERTDK-----EYIQIALF 1484
Qy 2751 -RWMTENSEYCKALMELEKFKSCDHC--KTSDCRNDYDENKCECKTRCQEKYKF 2806
Db 1485 KRWENF-----LEDYNKINDKISHCIKRGSGCINGCEN--SKLEKIEK-- 1532
Qy 2807 VLKWSLFDIOSNKKYELYE---OPITYKISTYDHOVNOFKLTFKSECVSFSFYH 2863
Db 1533 IAEMENI--KRFNDQYENKODPDYNVSI--LEELPKIYAVDQONVLIKLYE-- 1583
Qy 2864 ETSCLATKFNEDGSSNIRIYAFEEETPKSYKACSGTLPKSN--PLDNC--PTDQNKDG 2919
Db 1584 ENSKGCILISTONNKEN-----DAIDMLKLVKAKANCPGRSGEKQSD 1629
Qy 2920 CKELOTFTCSKNDYDNLDNWNAYVLNSSDNKGLVIPPBRRHLCSTRITAYNTRKGD 2979
Db 1630 CKEPPL-----PDEEDN-----PEENTLEPK--PCP--PTTOPPEK-- 1666
Qy 2980 KEILKKLLSASOGOLCOYKKEEELCFEAMKSYADYSITIKGTMDMPTLSLEK 3039
Db 1667 ----- 1680
Qy 3040 KIFETSEATENRKTWENNRROIWHAMLCGYKIATSKYTLDSGMCOLPDEETNOFLM 3099
Db 1681 K--BESSE----- 1693
Qy 3100 LIEMAKOACEKKHVSDSLTKCPRSNEDNFEASELLRQPCGCONDIRKYSILNLIKMT 3159
Db 1694 ----- 1717
Qy 3160 ENLINIKYKOLKDOSSGNIDNKPSSENVOSIYKSDOCALBLNDINEIYGTIKNNENEF 3219
Db 1718 ----- 1728
Qy 3220 KEVLAKLPLGYVEDETHKNHVLNIGNEEQYVRKALYFTTPHYDSFYAPLESTHR 3279
Db 1729 PPT----- 1749
Qy 3280 VAOYDPRNDILSSISVIVSALGIALHFMKKKFSY--DLRLIINIPQEGYGMPTLES 3338
Db 1750 ----- 1800
Qy 3339 KNRVTPYRSGPYKGYIYMEGDTSGDEKRYMDJSSDITSSSEBYEELDINDIYVGS 3398
Db 1801 SNVTIYVDRKYGKTYIYMEGDS--DEDKYAFMSDITDYVSSSEYEELDINDIYVGS 1858
Qy 3399 PKYTLIEVLEPSSKRDIPSD--DTPS-----NDTPRNFIDDENNELKHDVVSQYL 3449
Db 1859 PKYTLIEVLEPSSGNNTTASGKNTPSDIRNDIQDGLPSSKITDNEWNOJKEEFSIML 1918

[illegible]

RESULT 13

variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium falciparum)

Query Match	8.1%;	Score 1574;	DB 2;	Length 1711;
Best Local Similarity	17.6%;	Pred. No. 1.3e-65;		
Matches 635;	Conservative 259;	Mismatches 496;	Indels 2218;	Gaps 83

[illegible]

QY	580	NDNKRLLSEHNEFEELAWTYLLRDTIKRNDKLKJCINNTT--HCIDCANNCLCPBRWQ	638
Db	585	VYTP--ISYSFEFGSITIDMNEISEHREKRELCIDAKIGKCRKCKPCPCBYRWKEK	642
QY	639	KEEWNWSIKKLFYTKRKNIOOSYSNNINLEEGYFFVMDKLDEAKMELMENIKRKN	698
Db	643	KKEDEWKIKEFFRKQDUL--LKDIDMGAG-----ELLE-----	674
QY	699	EFSNLERNRDYLENAIELLDHLKETATICKDNNTVACETSHNATTNCPVPRGTOPT	758
Db	675	-----FYLEN--IFLEDMKN-----	687
QY	759	KNIEIAQYFKRSAYEEARNRGLHKLKGARHEGIVYKRGRRKDFKNLCRIKIKHSNRL	818
Db	688	-----	687
QY	819	GFSGNCPDCKGTGDGIOTRRVVGSTEWEDEBHRKDHEDVIMPPRRRHICTSNLEHLOT	878
Db	688	-----ANG-----DB-----	692
QY	879	DHPLNGNIYDVLVANNSPGLDVLBSAKYEAKIIRMKKEKNNLKGRPEVYDPRHOTTICRA	938
Db	693	-----KYLEKEKE-----	700
QY	939	IRYSFADIGDIINGRDLMERNGDWKLOGLHLETFVFNINHSILKKGKNDKYNDADARUKL	998
Db	701	-----LIGKRENEVOP-----	712
QY	999	RENWMEANRAKVEAMKCDIKLYLKDSGHOSTOSSYCYSDHTPLDDYIPOKLRMMTEMA	1056
Db	713	-----LTKTK-----TIDDFLEKEL-----	727
QY	1059	EWYCKVQKKEKDKLKECKECKDKDNQOGCTKESGTCCTACBACEYNDIIGLMEQWN	1118
Db	728	-----NAKKNCKVEKNPNBEPKQKAG-----	749
QY	1119	IISDYKKEHQAOMSVSNSNGIEASSTAKNHIIDRNVIEFLSELTYOONGSKSNKSGTSDS	1178
Db	750	-----DGAAPSDRPREDITNHH-----DGEHSDE-	773
QY	1179	AVIGNTTYENAVGYLHDTONFPODQOQONFQDEKSDGDKNEKYAFRDKROHDGACGK	1239
Db	774	-----DEEBEEDQPPRAEGTGESE--K	796
QY	1239	SGSKPTVOIKTKRKAEEKDETE-----CKTVNDIKENDGKQVDECHPRKNS	1286
Db	797	SESK-----EVEVEQOEPQDKDEKTVPTTPYVDCPVATALADGSLMA--CSLKYVT	850
QY	1287	NGYPRWOC--GNNINVEDRVCMPPRQKLCYHFLANDNEIKLLO--SQVNLKRAFIKS	1344
Db	851	GNKNGWRCIADSGTSGKDAICVPPRTIOELCYLT-----KELSDTYOKGILREAFIKT	904
QY	1342	AAAEFEFMYIYK--SKDGE--GNELD-----KELKEGKIPAPILRSMFTPDGDFLFG	1393
Db	905	AAOETIYLLMQYKKREDDKONERASTELDDIDDPQTLNGSELIDEPDKRMFTTFDQYDPLG	964
QY	1394	TDISKGEGSKLEQOJDSLFXNGDOKSPNG-----KTROEMWTEHSHEIMCALVYI	1445
Db	965	RYIG--NDLDRKVANNNTTAVFQNDH--IPNGQCTIDROBOEFMTGYDKIMKMLCALDBA	1020
QY	1450	GAKKDDFTENGYNNVNFPSD--KSTLEEFARKPOPLRMTLEMYDYCYTRQKYLKDQOE	1507
Db	1021	GGRK--TLETYINYNSVTFPNGHLGTIKLNEFASRPSFLRWATEGDFCBERITQOILKE	1075
QY	1508	KCKNSNDQLCDETCNKKCEBYVYKMKKKEMIPDQYKYVDERBDRKRRQHIGVWTDYT	1567
Db	1080	RC-----HYQGN	1087
QY	1568	GTNATDYINRKFASCGDCKPGSASVQYRNIQLEKQAYYADAKHGGCTFIENDOKYTRI	1622
Db	1088	G-----	1088

QY 1628 SSKDKCKGLYKANTGAIKMONKGNPNYNNIKELTEDVLPSSRLRICFHALDGNTPDE 1687
Db 1089 ----- 1088
QY 1688 VKDENGRLRLMEVAATEGYNLQGYKKEKKEKIKTSDAHKSYEVPPCSAMKYSFYDL 1747
Db 1089 ----- 1088
QY 1748 RDIIIGIDLEDEROKTEENIKRIFNKNGTSGKSDSTTGNGSTARKPFENNKECWA 1807
Db 1089 ----- 1093
QY 1808 NAMICGYKRRDNGSGNSARSDEDLKKGVSVDYPMGKNRDEGTAYOFLRPAEWG 1867
Db 1094 ----- 1093
QY 1868 EDFCKHEKELELVGACNDYTCGDNEDKRRKCTDCTOYKRFISHPKPOYEOKIKYGE 1927
Db 1094 ----- 1126
QY 1928 NKDKIYSEHPVAKADADAREYLDKOLKICENKSGDCEYCKMKDVSTORLTDGNSQMPA 1987
Db 1127 ----- 1126
QY 1988 SLDDPEKEVEGKCNQVPRGPPVRRETPSPRVSLSKATASKAKTAAPTOKRKYEN 2047
Db 1127 ----- 1126
QY 2048 LITTEKRAOTRTIRAAOQTRKRTSTATTESDVGTVKAILSNKPDPSRGIEGCPKTYGO 2107
Db 1127 ----- 1132
QY 2108 YPKWGIYKSKENENGICMPRRKKLCINNOYLYNETENKRDNDKEAFKCAIETQ 2167
Db 1133 ----- 1142
QY 2168 FLMLKTYIENPAENELONGTIPDEFKRMYYTYGDKDMFGTDSMDKLIITVNSVT 2227
Db 1143 ----- 1147
QY 2228 TLINENNNKKODKKDELRKTFWKKNNKFTWEGMITYGLTYHLNDENEKIRIDNOYOND 2287
Db 1148 ----- 1153
QY 2288 MTKLTPSLEEFYKRPQFLRWFTENAEFCNKRKQDLKLEAGCKYEGNGSDGTOCA 2347
Db 1154 ----- 1163
QY 2348 EACVYQNFYIKKWKTEYEROREKFKKDGKRYKDYSTERDIEKATCAHEYLNMKLKEL 2407
Db 1164 ----- 1166
QY 2408 CGNKDCSCNOKPSSQLPKTQOSOSDANDMPESDIYVPEFKN-CCEPPELSKSGSMITH 2466
Db 1167 ----- 1212
QY 2467 KKTTEPKTIPMNCVKAAYYSKEAENNDITLKERFPIESTKEKESKNSWTNNPCDPK 2526
Db 1213 ----- 1220
QY 2527 KPYAPDKYIGRRNPCENRENEKRVDEYEMCKYKSKFYQEKRRKVCVPRRHHMCLRNIDE 2586
Db 1221 ----- 1220
QY 2587 IKIBRLKDSNYLTKMVRRTAREGIDILIKNPNSENGCAMPICDTMKYSPADIGDIYRG 2646
Db 1221 ----- 1220
QY 2647 DMLRIGYLPPEVEIKLYKVEEYIYGKWRNNKNGKNTYDVQFERSAWANDNRDINRAMT 2706
Db 1221 ----- 1220
QY 2707 CKAPEDAKLFRKGRMDGFERITLLQDKGCHKDDPVDVDTIPQFRWMTSEYTCALME 2766

Db 1221 ----- 1220
QY 2767 ELEKFRKSCDHCKTSDRCNDYDENKCEQCKTRCOEYKNFVLKMSLPIQSNKKEIYE 2826
Db 1221 ----- 1220
QY 2827 QPIYTKISTYDVQNFVOKLKTFFKSECVSESESEYLETSCLYKNFNENDGSSNIRTYA 2886
Db 1221 ----- 1220
QY 2887 FEETPKSYKACSTLPSEKPNLNDPCPTDQNDGCKELQTFPCKSNNDYDNNLDMMNATLV 2946
Db 1221 ----- 1223
QY 2947 LNSDDNKGVLIPRRRLCTRPITAYNRKRGDEILKRLKLTSAFSOGQLLAGOKYKSEE 3006
Db 1224 ----- 1226
QY 3007 ELCFEAMKYSYADYSDILKGTDMMDTSLSEKIKKIFETSNEATENKRTWENNRQIWA 3066
Db 1227 ----- 1226
QY 3067 MLCGYKATSKVTLDEBGCQLPKDEETNOFLRWLLEWAKQACEKKHVSLSLTKCPRSN 3126
Db 1227 ----- 1229
QY 3127 EDNFEASELROPGCQNDIRKTYISLNLIKNTMENLKITKOLKDOSSGAINDKPSEENV 3186
Db 1230 ----- 1232
QY 3187 QSYTKSDSQCALBLNDINEIYGTGKNNNENNEFEVLKLYPGLYFEDETGHKNHYLDGN 3246
Db 1233 ----- 1236
QY 3247 IKKEQYVRKALYFTFPYHDSFYQAPLFTSHRYAQDPKNDILKSSISVIYALGLIA 3306
Db 1237 ----- 1263
QY 3307 LHFKKRKFSSVDLRLINIPQEGYGMPTLESKNRYIPYSGYKKGKTYIYMGDT---S 3363
Db 1264 YFYIAKKTKSTIDLKRYINIPKSDYDLPKLSPPRKYIPYSGYKRGKRYIYLEGDSGTDS 1323
QY 3364 GDEDKYMDLSSSDIT-SSESEYEELDINDIYVPGSHKTYLLEVLVLEPSKRDIPS--DD 3420
Db 1324 GYTDHY-----SDITSSESEYEELDINDIYAPBAPKTYLLEVLVLEPSGNTTASGNN 1377
QY 3421 TPS-----NDPRTNRFTDDENNELKHFVSOYLPNTEPN---NNYSADIPMNTPEPN 3470
Db 1378 TPSPDTONDIONDGIPSKITDNEMNWLKDEFISQYLQSEQPDVNDYSSGDIPLNTQPN 1437
QY 3471 TLVSDNPEEKPLTISHRDLYGKEISYINNSTWNNNDIPNANDYSGRIDGLINDSL 3530
Db 1438 TLVFDNPEKPLTISHRDLYGGEYSYVNMV-VNTNNDIPISGKNGTSGIDLINDSL 1496
QY 3531 VVALNLLTY 3538
Db 1497 NSNNVDLY 1504

RESULT 14
T18399
Variant-specific surface protein 1 homolog 3D7var1 - malaria parasite (Plasmodium fal
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18399
R:Rubio, J.P.; Thompson, J.K.; Cowman, A.F.
EMBO J. 15, 4069-4077, 1996
A:Title: The var genes of plasmodium falciparum are located in the subtelomeric regio
A:Reference number: Z18927; MUID:96324414; PMID:8670911
A:Accession: T18399
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-2042 <RUB>
A:Cross-References: EMBL:053324; NID:g1297090; PID:g1297091; PIDN:AAB09769.1
C:Genetics:
A:Map position: 12
A:Note: 3d7vari

Query Match 7.9% Score 1523.5; DB 2; Length 2042;
Best Local Similarity 19.3%; Pred. No. 3.7e-63;
Matches 680; Conservative 282; Mismatches 661; Indels 1895; Gaps 108;

QY 166 HDLGNVLTAKYGESEIYNNH-PKH-----GTSACTALARSFADIGIDVIGIDMEKPN 219
D 12 HNLIDVCLAAQYEGOSISGNHGKHOLSTYPSQJCTELARSPALIGDVRDLRYGN 71
QY 220 --VHREVTGLREFFKIHDMG---EDEYKNDYNPDGSGNYKLEAMNNVNRNWEA 273
D 72 NRENDLEKKLKYFKKIYDNLVEKKKEAETDYK--DDAPDFYQLREDWALNRQDWKA 130
QY 274 ITCDASYKSGYFMQSSNPPLFSPKCGHKQGVPTNLDVPOYLAWFDMGEFGRKN 333
D 131 ITCDAA-HDSYRRMGADGSTESAMQCRNAVADPTNFDPQYLKMFEMADFCRRK 189
QY 334 IKLKAYKDSGRND---KERLYCSHNGHDTTITMKKGIHLDNKCTDCKTCKVEYWL 389
D 190 HKLKDALQCRQDGTGKDR-YCDLNRVDOCKRTISAKHELVOGECKKCSVCFPGPW 248
QY 390 GNOQAFKQKEREKEIQ-----SYLSDNKFVNNINSEYK---QFYELKETQY 438
D 249 DNKOFEKQKNNYTNINKKHDETTKEISGNRRKRSLTKNKKYDEFEKIFKD-EX 307
QY 439 ATNDFPLNLLNEKGYCKG---GLPGKIDITFTNSADDKGTFYSEYCOVCPDGY----- 490
D 308 PDVDAKLDLISKTACESEPYPDEPRTISINEFN-YKNPDIFFSTHYCOACPMGGMWTL 366
QY 491 ----KCDGIKTHKSDNDRERNEDYKRPWGKPTNITVLVSGNOGDITOKLEFCN 545
D 367 MVAWQTORVYATIKLFKKNTLQI-----TTDPIPLPDTKSAIVKYNFCN 416
QY 546 SSNTYDKNNQKQECYKDEINIRKLEONTEINNDN-----PKITSFHNFE 593
D 417 SSDNNSDQINMCHYD-----ESKSGQDNQCVGTQWQNFKKQKVTSYNAFW 467
QY 594 LWTYLLRPTIKNDKLTCTCINNTTHCIDEARNCLCPDRWKKOKEEENSILKLTFR 653
D 468 KWSSEMLDQIKWRAELDKCLANDKKTGKKNYKNCILNFKLGNLKKKNGKQKHFN 527
QY 654 KNIQOSYYS---NINLFEGYFFKYMDKLKDE---AKWELMENIKRKNFESNLEN 706
D 528 KKIDRLVCLLKHVLLNDVFLQDMKAGDPQHIKAKIKELKKNDEKVNLSNMTI 587
QY 707 RQYLENAIELLDHLKETATICKDNNTNEACETSHNATTNFCVKKPRGCTQPTKNIK 766
D 588 FDF-----LQEBQDQOKCVSNP-EKCEETOKRPTDGA--PGA----- 625
QY 767 YFRSAVEEARNGLHKLKGAKAHEGIYKRGGRKDFDNLCRIMKHSNNLGSNPGD 826
D 626 -----GSP 629
QY 827 GKGTGDIOTRFVYGTWEVDEPHMRKDEVDLMPRRRHICTSNLHLOTDDHPLNGI 886
D 630 DTGTDNLE-----DIDSD--GEEDDV-----SHVDEEPEDNPVGE-- 665
QY 887 VDDLVNNSFLGDVLLSAKYEANKIIMYKKEKNLKGPKVEYTDKHOITJIRAIRYSADI 946
D 666 -----SSEKEQEVK---DTEAAYPKDQTPKEEVNPKIYEELF--- 703
QY 947 GDIIIRGLMERGDMVKLQGHLETFEGNIRKSLKSGNDKYNDADPKYLKLENNMEAN 1006
D 704 -----KSTKNFEA- 712
QY 1007 RAKWEAMKDIYLYLKDQSGHSTQSSYCGYSDHTPLDXYIPQKLMMTQWAMWYKQVOR 1066
D 713 -----CGLKYGKNYG-----MKC----- 725

QY 1067 KEYDKLKECKECKDKDNGQCGCTKESGCTCKCTEACNEYNIDITGLMKEQMIISDKYKE 1126
D 726 ----- 725
QY 1127 LHEQAQMSVNSGIEASSTAKNHIDRVLEFLSEIVQONGKSNKSGTSDESAIVIGNTT 1186
D 726 VHH-----TSDKQSEPTLARGHSHVANS----- 747
QY 1187 YENWGAYLHDTGNFDQSQSNEFCDEKSDGKDNKAYAFRDKPODHDGACGSKGSPTRV 1246
D 748 -----ADAPSGDK----- 756
QY 1247 QIKTKKAEEKTECTVNDILKENDKQVEDCHPKKNSNGYPMQCGNINLVEDPRVC 1306
D 757 -----DGAIC 761
QY 1307 MPRRQKLCVHFLANDNEIKKLOQVNLKRAFTKSAAEFFESWYK----- 1354
D 762 IPRRRRLYLHKT---EGVDTTDDKSLRKWFTESAVERFFLMDRKKLITPOSGSPLL 817
QY 1355 -----SKDEGNELDKELKGRIPAFRSMFYTGDRDFLFTGDISKHG--- 1401
D 818 GMITGVGVNDDENNPP-EKLLQKGEIPDGLRQMEFTYIGDYRDILFSGDKKKNQYSD 876
QY 1402 -----ESKLEQIDSLFKNGDQ-----KSPNGKTRQEWNTSHSHEITWAMLC 1444
D 877 IVSGDWIKERENTYIKKIASFFQNGKKEGTPIYKPN-VQTPQWKKDNKHHIMGMIC 935
QY 1445 ALKIKAKNDPFTENGYNNVKKFSDKSTLIEFAKRRQFLRWLTWDDCYRQYLYKD 1504
D 936 ALV----- 938
QY 1505 VOECKSNQDKCDTECKNKCEDYVYKMKKKKEMIPQDKYKDERKRPDRQHIGWY 1564
D 939 YEETSGSDEK-----KLEKDAYKKR----- 962
QY 1565 DYGTNATDYLANRFTASCQDKPQSGASVQORNIQLEKQAYADAKHCGCTKTENDKY 1624
D 963 --FTPGNPLPQPGTNGTSNEPIS-----QYQYD----- 990
QY 1625 TINSKDKCGLKVEANTGAIKQONKGPN---NYNNLKELTVEPLPSRRLRICFHALDG 1681
D 991 -----QVYLKEENNGMSTSPKSTSAAPSDNPTTLTQVLRPT----- 1029
QY 1682 NYTDPEYKDENGLRKRLMEVAATEGYNLQYKKEKEKEKIKTSDAHKYSYVPCSAM 1741
D 1030 ----- 1029
QY 1742 YSFYDLRDIILGIDNLEDEKQKTEBNLKITFNKNGTSVGKGSSTYGNPGSTARKFPWNE 1801
D 1030 ----- 1029
QY 1802 NKBCVNMAMICGYKRGDRDQNSGNSARSDEDLKCGSVPSDDYPMCKNNDGTAVOFLR 1861
D 1030 -----YFR 1032
QY 1862 WFAWGEDFCRKEKELEKLVGAC--NDYTC-GDNED--KRRK-----CTDA 1903
D 1033 YLEBMQGNFCRKKRRLAQIYEDCRGNDKYCSGDGECCEVVRQDSKISFNFCQGGRE 1992
QY 1904 CTQYKFFISEMKQYEQEQLKKGENDKIYSEHPVAKADADAREYLDQKLKICENKSGD 1963
D 1093 CRKYKWKIKRTREFEQKKAAYVDRDK----- 1120
QY 1964 CEYCKMKDVSQRLTDQNSQNPASLDDEPKVEYEGKCNQVPRGPPRVRRPETSPPVSLI 2023
D 1121 ----- 1120
QY 2024 SKATASKREAKTAPPTKQPKVENLTETEMRAQRTRRAAQTRKRSTATTTESDVGTW 2083
D 1121 ----- 1120

```

QY 2084 KALISNPKDSRGIEGNCPTTYGQYPRKGCIVGSKSENGICMPRRKLCIINNIOYLN 2143
Db 1121 ----- 1120
QY 2144 YETENKRDNDIKFAIFKCALETQFLWKYI IENPAENELONGTIPDEKRIMYTYGD 2203
Db 1121 YKTEKN ----- 1131
QY 2204 YKDMFSTDISNDKIIITVNSVTTLINENKKKODKKDELRKIFWERKKKFIWEGMI 2263
Db 1132 ----- 1131
QY 2264 YGLYHLTDENEKIKINDYQYNDMTKLPSPLEEFYKRPQFLMPTIEMAEFCNKRKEOL 2323
Db 1132 ----- 1133
QY 2324 LKLEAGCKEYECNSNDKGTQCEACACVYQNEIKRKWKTEYERQREKFKDKDKKYYKY 2383
Db 1134 ----- 1148
QY 2384 PSTERDIEK-ATCAHEYLNNKIKELCGNKCSCMOQPSOQPKTTOOSOSDANDPEST 2442
Db 1149 ASTESPLEKIGSCSK ----- 1171
QY 2443 DYVEEFNKCEPELSKSGSMITTKITEPKIPMNCVEKAAAYLSKEANNDITLKEKF 2502
Db 1172 ----- 1183
QY 2503 IPIESTKEKSKNSWTNNPCDPKPPADPKYIGRRNPCENRENFKYDYEMKCYKNSK 2562
Db 1184 ----- 1202
QY 2563 FVOEKRY-CVPRRREHMLRNDELKIRLDSNTLKMVRRRTARNESIDITIKPNSEN 2621
Db 1203 CSGGCKKVC ----- 1231
QY 2622 GCAMNFICDTMYSFADLDIVRGDMLRIGYLPVEIKLYKVEEYIYKWRKNNKGN 2681
Db 1232 GYFMT-VSDNSDKFE ----- 1259
QY 2682 KYNDQVTFSSAMWANDNRKDIWKAMCKAPEDAKLFRKMGDFERTITLIDCKGHRDPP 2741
Db 1260 ----- 1287
QY 2742 VDDYIPORFRMTSEYCYKALMELEKFKSCDHCKTSDR--CKNDYDNCKEQCKT 2798
Db 1288 GKELYQIR-ALKRWVEYF-----LEDYNNKIKKISHCARKGKCTICIND----- 1331
QY 2799 RQGEKKNVYKKSFLDIQSNKYKELYQIPYTKISTYDHYQNFQYKLTFRKSECSVESF 2858
Db 1332 -CVE-----KW--INIKKKDMETIRER--YVKQYTTGH----- 1359
QY 2859 SEYLAHETSKCLNFKNENDGSSNITFAFEELPKRYKACSGTLPKSNPLNCPDQNKD 2918
Db 1360 ----- 1384
QY 2919 GCKELQTFPCSKNDYDNMLNMNAYLVLNSSDDKGGVLLIPRRRLHCTRTITAVNRKG 2978
Db 1385 -CGDLDFKQ-----NSTD-----CT-----VYRSS 1403
QY 2979 DKEILKRLLSAFSOGOLLQOKYKSEELCFEAKKYSYADYSDI IKGTDMMDTSLSEKI 3038
Db 1404 ENGVYNNK-----DIVECLL----- 1418
QY 3039 KKIETSEATENRRTWMENNRQIWHAMLCGYKIATSKVTLDEGWCQLPRDEETNOFLR 3098
Db 1419 -----ENLKT-----KAKTC-----PNO-- 1431
QY 3099 WLITMAQACEKKKRVSDSLKTKCPRSNEDNEFEASELLRPGCONDIRKYISLNLIKNT 3158
Db 1432 --VANGENOTCJSLPHVEDD-----DDEBPLEETE-----ENT 1461
QY 3159 MENINIKYKQAKDQSSGNDKPKSEENVQSIKSKDSOCALELNDINEIYTGKNNENNE 3218

```

```

Db 1462 VEPNIPCFQDLPKPP-----QPEDGCKQASPAPE-----GTENO----- 1498
QY 3219 FKEVLLKLYPGLYVEVDETHKNHVLIDGNKEEQ-----TVPRKALYFTTPVDSFYQAP 3273
Db 1499 -----PVY-----IKPEEAPAPRTPRRK-----PQEPQYLP 1530
QY 3274 LFSTRVAQYDPRKNDILKSSISVIYSAIGLIALFMKKFRSSV-DLRLILNTPQEGY 3332
Db 1531 AL-----KNMLSTIMWSIGIGPAFTYPLKKSXSVGNFQIOLQIPKSDP 1580
QY 3333 MPTLESKNRYIPYRSGPYKGYIYMEGDT--SGDEKYMWDJSSSDIT--SSSEYEEL 3388
Db 1581 IPTKLSPNRYIYPTSGKRAKRYIYLEDGSGDGYTHY-----SDITSSSEYEEM 1634
QY 3389 LINDIYVGSPPKYLIEVLEP-----SKRIDPSD--DTPSNDTPRTNR 3431
Db 1635 DINDIYVPTPKYKYLIEVLEPSSGNNTTAGSNNTTAGKMTPSDQNDIPSQTP-NNK 1693
QY 3432 FIDDEMNELKDFVSOYLPTNEPN--NNYKSADIPMTPEPNTLYSDNPEKPTIISHD 3488
Db 1694 LFDNEMNLLKDEFISNMLQSQPKDVPNDYSSGDIPTFTQSTLYFPDNOEKPTITSIDH 1753
QY 3489 RDLTYGKEISYNNKSTNTNNDIPMANRDSYRGIDL 3526
Db 1754 RDLTYGEEYNTYNNKSTNSMDIPISGKNDYSGIDL 1791

RESULT 15
T18396
E:lytocyte membrane protein variant 2 - malaria parasite (Plasmodium falciparum) (fr
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18396
R:Baruch, D.I.; Pasloske, B.L.; Singh, H.B.; Bi, X.; Ma, X.C.; Feldman, M.; Taraschi,
A>Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen a
A:Reference number: Z18925; M01D:9530812; PMID:7541722
A:Accession: T18396
A>Status: preliminary; translated from GB/EMBL/DBD/J
A:Molecule type: mRNA
A:Residues: 1-1729 <BAR>
A:Cross-references: EMBL:U27339; NID:9914920; PID:9914921; PIDN:AAA89134.1
C:Genetics:
A:Gene: EMP1
A:Note: var-2

Query Match
Best Local Similarity 22.3%; Score 1249; DB 2; Length 1729;
Matches 503; Conservative 273; Mismatches 632; Indels 846; Gaps 88;

QY 33 KSANVLEARYAKNI-----RHPSKYAKEHVDSLKGLDTKAEEFGSPYVNNHNYTY 86
Db 13 KDAKHALDRIGEYVYKKEVNDAAEKTK--ALKGNLDQAKGIGELASS-----PN 60
QY 87 PCNLDIKHEATN-LRYDDVNLHPCHGREGONRFDEDESEGC-NKIRNYRKNDAL-ACAP 143
Db 61 PKRLVEDYVNNRLK-----RKRYPCANRQYVRFSDYGGCGCTFRIRKIDSNMNSIGACAP 116
QY 144 PRRHMDKRLLEALNDINQNIHDLGNVLYTPAKYGESIVNNH-PHKT-----SDACT 197
Db 117 YRLHLCDIYLEMGGKTSYTK-HDLLLDVCMARKYSGDISIKHYKHETLNPDYKSQLCT 175
QY 198 ALARSEADIGDIYRGIDM-----KPNVHDKVETGLRVEVFKIHGDMDEYKKNYNDPG 251
Db 176 ILARSEADIGDIYRGIDYLYGYDKREKDERKLENNLILFIKKIHENLGTQAKKHKKD 235
QY 252 SGNVYKLEAMNMYNNKRWAEATCDASYSKGYFMQSENTPLFSNP-KGCHKQCK----- 306
Db 236 EENYVYDLEWMTANSTWAKATCHAGSSDYFRFTKCSGEMTDDCKCRCKDEKKNFTN 295
QY 307 -VFTNLDYVQYLRWDEWGEERCKRNKIKLKYVDSGRNDK--BRVCSHGHCCTTI 363
Db 307 -VFTNLDYVQYLRWDEWGEERCKRNKIKLKYVDSGRNDK--BRVCSHGHCCTTI 363

```

Db 296 EYPTTYDYVQYLRMEFEWAEDEFCRRKKRIENAIKNGEKGNER-YCDLNGNCEETA 354
 QY 364 MKKGIHLNDKCTCSTCKVEEVLGNQOEAFKQKEREKEIO---SYLSNDKRYN 419
 Db 355 RGAEIFVKGDDCHKCSVACDRFVWIDNQKREDDKOKKKYDEIHGHTITITGKGN 414
 QY 420 NINSEYKQPYELKCTOATNDTFLNLNBEKCYK--GGLGEKQITITNSADKGIY 477
 Db 415 NL---VGFHFYKILKK-YYPYDRKSLQKLNDEAIKCPNVGNEKSTDFNNEVTTFS 470
 QY 478 RSEYCOVPCDCGVKOD---GKRYTHKS--DNDBERNYNDYKRPWGVKPLNITVLSGNE 532
 Db 471 HTTYCAGCWCAGQKNGKNGKAKKSCAKKERIFNKE-----NSTDIKILPEKG 523
 QY 533 QGDITQKLENFCSSTNYKDNQKQWECYKDEINRCLEONTLEIND-----581
 Db 524 RSKTELEKTCFCGDKQIK---NDIKCHYDNG---TDQDITDSNOCVLGDWGNLTK 575
 QY 582 NKRIIFHNFFELMYTYLLRDTIKMNDKLTCTN-NTTHCIDECARNCLCFDPRWYKOE 640
 Db 576 EDKIMSYNAFPMWVMDLIDISIKWDEHGRICNDKRGTCIKGCKKICQKQWYEOBK 635
 QY 641 EEWNSIKLFTKTKKNIQOSYVSNINLFGYEFKVDKLDKDEAKWEIMENIKRKKNE 700
 Db 636 TEMGKIKDHFRRKOKDIPKDW-----655
 QY 701 SMLENNRDYLENAI--ELLDLKETATICKDNNTNEACETSHNATNCPYKRGSTOPT 758
 Db 656 ---THDDFLQTLMLKDLLEITIDTY---GDANE-----683
 QY 759 KNIKELAQYFKRSAYEABANRGLKLGKAHBEITYKRGRRDFKCNLCRIMIKHSNRL 818
 Db 684 --IKRI---EALLBQA---GVGGIDFALAGLYKGFVAE--KDTITDKLLQHOOK-- 729
 QY 819 GFGNGCDCKGTGDIQIOTRFVVTGEMVEDPEHRKDEHVEDYPPRRHICTSNLEHLQTD 878
 Db 730 ---BADCKLTHTDTCPPQ-----E 747
 QY 879 DHPNGNIYDVLNNSFLDVLISAKYEANKIIRMYKERNNLKGPEVTPDKHQTTICRA 938
 Db 748 DRSV-----ARSESATV-----PSPPADPK-----767
 QY 939 IKSFPADIGDIIRGRDLBERNGDMVLOGLHLEVPNGNIHKSILKGNKNDKYNDAPYTKL 998
 Db 768 ---ATEEVANASSDEDEDFEEB-----787
 QY 999 RENWNEANRAKYWEAMKCDIKYLKDKSGHSTQSYCGYSDHTPLDYIPIQKLRMTENA 1058
 Db 788 ---787
 QY 1059 EWCYKQKEYDKLEKCKECKDKDNGOGCTKESGTCTKTEACNEYNDIIGLMEOWN 1118
 Db 788 ---EEEBEDEGE-----796
 QY 1119 IISDKYKELHEQAOQSVNSGIEASTAKNHIDRWITELSELXYONGCKSNKSGTSDS 1178
 Db 797 ---EEAE-----800
 QY 1179 AVIGTITTEYENGAYLHDITGNFDDCOSONEFCDEKSDGKNERYAFRDRQDHDGACGK 1238
 Db 801 ---EVOBEKTDSESTEVA-----816
 QY 1239 SSGSKPTRVQIKYKKAEEKDTE-CKTVNDILKENDKKOVEDCHPRKNSNGYPMQCGNT 1297
 Db 817 -PSPPTQDGVPAQEDDVKCSIVDKALK---GKLD-DACITLYGKTAFTSMKC---868
 QY 1298 NLVEDPR-----VCMPPRQKLCYHFL-----1319
 Db 869 -IPSDTKSVATTGSDTGGSGSICVPRRRRLYGLKLDWAGGETTEAKSOETISGQKTPS 927
 QY 1320 ANDN-----EIKLOSQVNLKEAFIKSAAEFTFSWYYSK-----1356
 Db 928 GNSSPSEKLPQGPTEPTEJTKETPESSLHAFAVSPPLRLRFLPMHKKREOMKQHGAGATC 987

QY 1357 -----DGEINLDEKLEKGIKIPAPILRSMFYTEGDRDLEFGTD-----ISK 1398
 Db 988 LQLPGVYDSDPDPOQTQKRGNIIPNDLROMFYTLGDRDLCIGDRDRIYGVDTIVSITE 1047
 QY 1399 GHGEGSKLEQIDSLFKNDOKSPNCK-----TROEWTEHSHETWA 1441
 Db 1048 GESTKRIKISILIEGLKQOTVSPSPRDTSSRTVPHPQTSVEKTPPOOTWEANGPPIWNG 1107
 QY 1442 MLCALV-----KI-----GAK---KDDPTENYGNVNVFSDK 1470
 Db 1108 MICALITYEDSGAIGOPPOVEDADKLEKLPNTANGIKWYLEKEDTSSAMPSSSSSG 1167
 QY 1471 ST-----TLEEFAPKPOELRWLTEMWYDYCYTRORYLNDVOEKCSND-----1513
 Db 1168 SNDPINTPLTEVEIETFEFRYLHEWGNQFCKERMRLKOIYNECKVGENGVRGRKORT 1227
 QY 1514 -QLKC-----DT-----EONKCEYVYKMKKK-EMLPQDK-----1543
 Db 1228 PQSCYGEDECDQLSKYSYDTVADLECPKCAKRWYKWKWIEKKKDETEQEKAPFKOD 1287
 QY 1544 YKDERDKRRFRD-----OHIGVATDYGTATDYL-----NRKFTAS 1582
 Db 1288 YVYNGNKKGGGDMGFCITILKSLSDAQFLEKLSCKKDSNGDKLNFSPNETFVA 1347
 QY 1583 CGDKPGSASVYONIQLEKQAYYADKHQCTKFLENDKYTNISSKDKGLVYEANT 1642
 Db 1348 TNCKPCSEFKIDCKENGKCKNGGCTNENCTGTFT-----TSENFKQKG--QTAKE 1397
 QY 1643 GAIKQONKGPNNYNNILKELTEV-LFPSSRL-----RICFHAL-----DGWYTPPEK 1689
 Db 1398 FVNRVSDNNNGFDLNEACQNGAIFKSIKRDWEGCKVCGEVCJPEKNGNYTSGENN 1457
 QY 1690 DE---NGLRKRLEVAATEGYMLQYKREKKEKEXIKTSDAH-KYSEYPPC-----1737
 Db 1458 DOIITIRGLVAHVQ--NELDYN-----KIKKIKSHCKNSSGTYCTIKKVCQD 1504
 QY 1738 --SAMKYSFYDLIDILIGIDNLEDEKQTEENLKI-----PNKN---GTSVGKG 1782
 Db 1505 WISTKRTWNTNIK-ILLN---EQYKNDPDYVNYKTILQDLOQIDFNKAIKPGCT-LTKF 1558
 QY 1783 SDSTGNPGSTARKFTWNEKCEVWNAAMICGYRGR---DDGSGSASDEDLKCK--G 1837
 Db 1559 EDSGGINGAESSEKKNHGE---YDAIDCMNLRQDKIDDCRKNHAONGENGQACERH 1613
 QY 1838 SVPSDDDYPMGKNRDEGTAYQFLRMFAWGEDEFC-KHKEKELEKLVGACNDYCGDNEDK 1896
 Db 1614 SAPDEDDDEALIEENPYTO-----PNICPAPPEPKAEKGG-----CEPAEKK 1656
 QY 1897 RKKCTDAGTQYKKFISEMKRQYERKQIKKYGENKDKITYSEHPV--AKDADAREYLDKOLK 1954
 Db 1657 EKV-----EKEKEKYNTVAKPTEKEAAGDPAPADSEENPE-----1694
 QY 1955 KICENKSGDCYKCMKDYSTQRLTDGNSQNMPS 1988
 Db 1695 ---EERAPPEVETKDKAPVKPT-----PAS 1718

Search completed: April 28, 2003, 10:32:38
 Job time : 211.847 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 108.055 Seconds
(without alignments)
4367.892 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407
Sequence: 1 MFSCSKYFIIMGNASSLE.....IDLINDSLVNLILYMKY 3542

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

al number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDs2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19407	100.0	3542	22	AA62142
2	3489	18.0	3060	18	AAW22475
3	3489	18.0	3060	21	AAV77905
4	3108.5	16.0	2913	17	AAW00384
5	2972.5	15.3	2710	18	AAW22482
6	2972.5	15.3	2710	21	AAV77904
7	2972	15.3	2703	16	AAW70236
8	1749	9.0	2197	21	AAW18352
9	1743	9.0	2182	18	AAW22476
10	1743	9.0	2182	21	AAV77906

11	1681.5	8.7	2228	20	AAW93944	P. falciparum PFEW
12	1244.5	6.4	1726	17	AAW00385	Truncated Plasmodi
13	1045.5	5.4	1700	21	AAW18144	Plasmodium falcipa
14	1032.5	5.3	700	16	AAW70235	P. falciparum EBL-
15	1032.5	5.3	700	18	AAW22481	Plasmodium ebl-2.
16	1032.5	5.3	700	21	AAV77903	P. falciparum ebl-
17	827	4.3	407	21	AAW62148	P. falciparum varc
18	811.5	4.2	431	21	AAW18350	Plasmodium falcipa
19	665.5	3.4	1086	23	AAW76760	Plasmodium falcipa
20	655.5	3.4	294	22	AAW62147	P. falciparum varc
21	636.5	3.3	1604	16	AAW70105	TNF-R-EBA 175 fusi
22	633	3.3	921	18	AAW22480	Plasmodium E31a.
23	633	3.3	921	21	AAV77902	P. falciparum ebl-
24	631.5	3.3	793	16	AAW70234	P. falciparum E31a
25	631.5	3.3	1786	14	AAW41043	CD4-EBAl75 fusion
26	617	3.2	1435	16	AAW70232	P. falciparum SABP
27	617	3.2	1435	18	AAW22477	Silicic acid bindin
28	617	3.2	1435	21	AAV77900	P. falciparum SABP
29	609	3.1	311	22	AAW62150	P. falciparum varc
30	599.5	3.1	1421	23	AAW76764	Plasmodium falcipa
31	594.5	3.1	1501	23	AAW76762	Plasmodium falcipa
32	591.5	3.0	1143	23	AAW76759	Plasmodium falcipa
33	572	2.9	308	22	AAW62151	P. falciparum varc
34	550.5	2.8	351	22	AAW62149	P. falciparum varc
35	529	2.7	616	23	AAW50533	Unidentified amino
36	523.5	2.7	440	21	AAW18146	Plasmodium falcipa
37	495.5	2.6	445	22	AAW66344	Malarial parasite
38	431	2.2	749	16	AAW70233	P. falciparum EBL-
39	431	2.2	749	18	AAW22479	Plasmodium ebl-1.
40	431	2.2	749	21	AAV77901	P. falciparum ebl-
41	421.5	2.2	1979	21	AAW18171	Plasmodium falcipa
42	398.5	2.1	5024	22	AAW62835	S. epidermidis ope
43	396	2.0	10182	23	ABP38314	Staphylococcus epi
44	384.5	2.0	2573	21	AAW18234	Plasmodium falcipa
45	382.5	2.0	972	23	AAW76761	Plasmodium falcipa

ALIGNMENTS

RESULT 1	AA62142	standard; Protein; 3542 AA.
ID	AA62142	
XX	AA62142	
AC	AA62142	
XX		
DT	29-MAY-2001	(first entry)
XX		
DE	P. falciparum FCR3	varcSA protein.
XX		
KW	FCR3	varcSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
KW	erythrocyte membrane protein 1; parasitized red blood cell; PRBC;	
KW	malaria; protozoacide.	
XX		
OS	Plasmodium falciparum.	
XX		
PN	W0200116326-A2.	
XX		
PD	08-MAR-2001.	
XX		
PF	01-SEP-2000; 2000MO-US24195.	
XX		
PR	01-SEP-1999; 99US-0152023.	
XX		
PA	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PI	Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;	
XX	Gysin J, Pouvelle B, Fujii N, Smith J;	
XX		
DR	WPI; 2001-235109/24.	
XX		
DR	N-PSDB; AAF57301.	
XX		
PT	Novel FCR3	varcSA protein, useful for modulating parasitized red blood

PT cell binding, sequestration and onset of maternal malaria -
 XX Claim 12; Page 63-71; 78pp; English.
 CC The invention relates to a P. falciparum FC33, varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PEM1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient identified at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents the P. falciparum FC33, varCSA protein.
 XX Sequence 3542 AA:

Query Match 100.0%; Score 19407; DB 22; Length 3542;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3542; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MGPSCKYFLIKGNNAASSLSEDAKSPITIKSHKSAENVLEKAKNTRHPSKAKEHVDL 60
 1 MGPSCKYFLIKGNNAASSLSEDAKSPITIKSHKSAENVLEKAKNTRHPSKAKEHVDL 60
 61 KGDITKAEPFGSPVPVKNHNYYPYPCNIDKHEHTNLRYDVNLRHPCGSEONFEDD 120
 61 KGDITKAEPFGSPVPVKNHNYYPYPCNIDKHEHTNLRYDVNLRHPCGSEONFEDD 120
 121 ESEEGCNKIKNTKKNDAIADPSSRNQDKNLPLANDINTONTIHDLGNYLVAKYEG 180
 121 ESEEGCNKIKNTKKNDAIADPSSRNQDKNLPLANDINTONTIHDLGNYLVAKYEG 180
 121 ESEEGCNKIKNTKKNDAIADPSSRNQDKNLPLANDINTONTIHDLGNYLVAKYEG 180
 121 ESEEGCNKIKNTKKNDAIADPSSRNQDKNLPLANDINTONTIHDLGNYLVAKYEG 180
 181 ESIIVNHHPKSTSDACTALARSFADIGDLYRGIDMEKPVHDKVEGLRBYAKKHIDGME 240
 181 ESIIVNHHPKSTSDACTALARSFADIGDLYRGIDMEKPVHDKVEGLRBYAKKHIDGME 240
 181 ESIIVNHHPKSTSDACTALARSFADIGDLYRGIDMEKPVHDKVEGLRBYAKKHIDGME 240
 181 ESIIVNHHPKSTSDACTALARSFADIGDLYRGIDMEKPVHDKVEGLRBYAKKHIDGME 240
 241 DEKANDYNPDSGNYKYLEEAWNVNBNKWEALITODASKSGYEMOSESNPLFNSPKC 300
 241 DEKANDYNPDSGNYKYLEEAWNVNBNKWEALITODASKSGYEMOSESNPLFNSPKC 300
 241 DEKANDYNPDSGNYKYLEEAWNVNBNKWEALITODASKSGYEMOSESNPLFNSPKC 300
 241 DEKANDYNPDSGNYKYLEEAWNVNBNKWEALITODASKSGYEMOSESNPLFNSPKC 300
 301 GHKOKRVPMDYPOYLRWPEDEMEEFCKRKNITLKYKDSGRNDKERITCSHNGHDT 360
 301 GHKOKRVPMDYPOYLRWPEDEMEEFCKRKNITLKYKDSGRNDKERITCSHNGHDT 360
 301 GHKOKRVPMDYPOYLRWPEDEMEEFCKRKNITLKYKDSGRNDKERITCSHNGHDT 360
 301 GHKOKRVPMDYPOYLRWPEDEMEEFCKRKNITLKYKDSGRNDKERITCSHNGHDT 360
 361 TTIIMKRGILHLDNKCTDSTCKCIVEVWLGNOQEAFFKOKKEKEEIOSTYSNDKFPVNN 420
 361 TTIIMKRGILHLDNKCTDSTCKCIVEVWLGNOQEAFFKOKKEKEEIOSTYSNDKFPVNN 420
 361 TTIIMKRGILHLDNKCTDSTCKCIVEVWLGNOQEAFFKOKKEKEEIOSTYSNDKFPVNN 420
 361 TTIIMKRGILHLDNKCTDSTCKCIVEVWLGNOQEAFFKOKKEKEEIOSTYSNDKFPVNN 420
 421 INSEYKQFYEKLEKETOYATNDTFLNLNEGKYCKGLPGEKDIPTFNSADDKGIFYRSE 480
 421 INSEYKQFYEKLEKETOYATNDTFLNLNEGKYCKGLPGEKDIPTFNSADDKGIFYRSE 480
 421 INSEYKQFYEKLEKETOYATNDTFLNLNEGKYCKGLPGEKDIPTFNSADDKGIFYRSE 480
 421 INSEYKQFYEKLEKETOYATNDTFLNLNEGKYCKGLPGEKDIPTFNSADDKGIFYRSE 480
 481 YQVCPDGCYKCDGKIKYTHKSDNDRERYNNEDYKPMGVKPTNTTVLYSGNEQGDITOKL 540
 481 YQVCPDGCYKCDGKIKYTHKSDNDRERYNNEDYKPMGVKPTNTTVLYSGNEQGDITOKL 540
 481 YQVCPDGCYKCDGKIKYTHKSDNDRERYNNEDYKPMGVKPTNTTVLYSGNEQGDITOKL 540
 481 YQVCPDGCYKCDGKIKYTHKSDNDRERYNNEDYKPMGVKPTNTTVLYSGNEQGDITOKL 540
 541 ENFONSTNYVDKNNOKECYKDEINIRCKLEONTEINNDPKIISFHNFEFLMAYTLL 600
 541 ENFONSTNYVDKNNOKECYKDEINIRCKLEONTEINNDPKIISFHNFEFLMAYTLL 600
 541 ENFONSTNYVDKNNOKECYKDEINIRCKLEONTEINNDPKIISFHNFEFLMAYTLL 600
 541 ENFONSTNYVDKNNOKECYKDEINIRCKLEONTEINNDPKIISFHNFEFLMAYTLL 600
 601 RDTIKMWDKLTCTINNTHHIDICNRCCLCFDPMWVAKOKEEEMNSIKKLFYKKKNIOOSY 660
 601 RDTIKMWDKLTCTINNTHHIDICNRCCLCFDPMWVAKOKEEEMNSIKKLFYKKKNIOOSY 660
 601 RDTIKMWDKLTCTINNTHHIDICNRCCLCFDPMWVAKOKEEEMNSIKKLFYKKKNIOOSY 660
 601 RDTIKMWDKLTCTINNTHHIDICNRCCLCFDPMWVAKOKEEEMNSIKKLFYKKKNIOOSY 660
 661 YSNINNLFEGEFFKVMKLDKDEAKKKEIMENIRKKKNESFNLNNNDYLENAETILLDH 720
 661 YSNINNLFEGEFFKVMKLDKDEAKKKEIMENIRKKKNESFNLNNNDYLENAETILLDH 720
 661 YSNINNLFEGEFFKVMKLDKDEAKKKEIMENIRKKKNESFNLNNNDYLENAETILLDH 720
 661 YSNINNLFEGEFFKVMKLDKDEAKKKEIMENIRKKKNESFNLNNNDYLENAETILLDH 720
 721 LKEPATICKDNNTNACETSINATTPCVAKPRGSGTOPTKNIKETAOYFKRSAYEARNRG 780
 721 LKEPATICKDNNTNACETSINATTPCVAKPRGSGTOPTKNIKETAOYFKRSAYEARNRG 780
 721 LKEPATICKDNNTNACETSINATTPCVAKPRGSGTOPTKNIKETAOYFKRSAYEARNRG 780
 721 LKEPATICKDNNTNACETSINATTPCVAKPRGSGTOPTKNIKETAOYFKRSAYEARNRG 780
 781 LHKTKGAHEGITYKRGRRKDFKNLCRIMIKHSNRNLGSGNCGDKGIGDGIQTREYV 840
 781 LHKTKGAHEGITYKRGRRKDFKNLCRIMIKHSNRNLGSGNCGDKGIGDGIQTREYV 840
 781 LHKTKGAHEGITYKRGRRKDFKNLCRIMIKHSNRNLGSGNCGDKGIGDGIQTREYV 840
 781 LHKTKGAHEGITYKRGRRKDFKNLCRIMIKHSNRNLGSGNCGDKGIGDGIQTREYV 840

Db 781 LHKTKGAHEGITYKRGRRKDFKNLCRIMIKHSNRNLGSGNCGDKGIGDGIQTREYV 840
 QY 841 GTEWEVDPEHMRKDHEDVIMPPRRRHICTSNLEHLQTDHPHNGNIVDLYVNSSELDVYL 900
 Db 841 GTEWEVDPEHMRKDHEDVIMPPRRRHICTSNLEHLQTDHPHNGNIVDLYVNSSELDVYL 900
 QY 901 LSAKYANKITIMYKKNLKGPKVETDPKHOTTCORALRYSPADIGDIRDRDMERNNG 960
 Db 901 LSAKYANKITIMYKKNLKGPKVETDPKHOTTCORALRYSPADIGDIRDRDMERNNG 960
 QY 961 DMVKLOGLHLETFVGNHKSILKSGKNDKYNDAPKYLKRENNWEANRAKVMAMCDIKY 1020
 Db 961 DMVKLOGLHLETFVGNHKSILKSGKNDKYNDAPKYLKRENNWEANRAKVMAMCDIKY 1020
 QY 1021 LMDKSGHSTOSSYSGYSHPHTLDIYIQOKLRMWTENAEWTCYKOKKEYDKLKECKECK 1080
 Db 1021 LMDKSGHSTOSSYSGYSHPHTLDIYIQOKLRMWTENAEWTCYKOKKEYDKLKECKECK 1080
 QY 1081 DKDNGOGCTKESGCTCTEACNEYNDIIGLAKROWNISDKYKELHEQOMSVNSNGI 1140
 Db 1081 DKDNGOGCTKESGCTCTEACNEYNDIIGLAKROWNISDKYKELHEQOMSVNSNGI 1140
 QY 1141 EASSTAKNHDNRVLEFLSELYOONGSKNSGTSDESAYVIGTNTYENAGAYLHDITGNF 1200
 Db 1141 EASSTAKNHDNRVLEFLSELYOONGSKNSGTSDESAYVIGTNTYENAGAYLHDITGNF 1200
 QY 1201 DDCGQONFCEBESDKGKNEKTAFAEDKPODDHAGCGGSSGKPRVQIKTKKAEEKTE 1260
 Db 1201 DDCGQONFCEBESDKGKNEKTAFAEDKPODDHAGCGGSSGKPRVQIKTKKAEEKTE 1260
 QY 1261 CTVNDILKENDGKQVDECHPRKNSNGYPMOGGNILMYEDPRVCMPPRRQKLCVHFLA 1320
 Db 1261 CTVNDILKENDGKQVDECHPRKNSNGYPMOGGNILMYEDPRVCMPPRRQKLCVHFLA 1320
 QY 1321 NDNEIKKLOSOVNLKEAFIKSAAAFPPSWYIYYSKDEGEVLDKELKEGKIPPAFLRSM 1380
 Db 1321 NDNEIKKLOSOVNLKEAFIKSAAAFPPSWYIYYSKDEGEVLDKELKEGKIPPAFLRSM 1380
 QY 1381 FYTFGVDYDFEFGTIDSKGHEGSKLEQIDSLFKNGDOKSPNGKTOEWMTESHSIME 1440
 Db 1381 FYTFGVDYDFEFGTIDSKGHEGSKLEQIDSLFKNGDOKSPNGKTOEWMTESHSIME 1440
 QY 1441 AMLCALYKIGAKKDDFEBNGYNNVAFSDKSTLLEBAKRPQFLRWLLEWYDYCYTROK 1500
 Db 1441 AMLCALYKIGAKKDDFEBNGYNNVAFSDKSTLLEBAKRPQFLRWLLEWYDYCYTROK 1500
 QY 1501 YLKDVOEKCKSNDOJLKCDETCNKKCEDYVYMKKKKEMIPQDKYKDERKKRRDRHGIG 1560
 Db 1501 YLKDVOEKCKSNDOJLKCDETCNKKCEDYVYMKKKKEMIPQDKYKDERKKRRDRHGIG 1560
 QY 1561 VMATDYTGYNATDYLNKFTASCGDKPGSASVVOYRNIQLEKQAYYADADHCGCTFIEN 1620
 Db 1561 VMATDYTGYNATDYLNKFTASCGDKPGSASVVOYRNIQLEKQAYYADADHCGCTFIEN 1620
 QY 1621 DDKYTNISSDKCKGLYKEANTGAIKQONKGPNNYNNLAKELTEDVLPSPRRILICFHALD 1680
 Db 1621 DDKYTNISSDKCKGLYKEANTGAIKQONKGPNNYNNLAKELTEDVLPSPRRILICFHALD 1680
 QY 1681 GNTYDPEVKNENGLRKRLMEVAATEGYNLGOYKKEKKEKIKTSDAHKSYVPPCSAM 1740
 Db 1681 GNTYDPEVKNENGLRKRLMEVAATEGYNLGOYKKEKKEKIKTSDAHKSYVPPCSAM 1740
 QY 1741 KYSEFYDLRDIILIDNLEDEKOKTEENLKKIFNKNGSTSVAGKSDSTYGNPGSTARFFPN 1800
 Db 1741 KYSEFYDLRDIILIDNLEDEKOKTEENLKKIFNKNGSTSVAGKSDSTYGNPGSTARFFPN 1800
 QY 1801 ENKECYVNMAMICGKRGDRDGSNGSARSDEDLKKGVSVPDDDYPMGNRRDGTAYOFL 1860
 Db 1801 ENKECYVNMAMICGKRGDRDGSNGSARSDEDLKKGVSVPDDDYPMGNRRDGTAYOFL 1860
 QY 1861 RWFPEWGEDFCHEKELEKLVACNDYTCGDNEDKRRKCTDACYKKFISPMKPOYER 1920
 Db 1861 RWFPEWGEDFCHEKELEKLVACNDYTCGDNEDKRRKCTDACYKKFISPMKPOYER 1920

QY 1921 QIKKYGENDKITYSEHPVAKDADADAREYLDKOLKICENKSGCEYCKAKDVSTORLTG 1980
 DB 1921 QIKKYEENDKITYSEHPVAKDADAREYLDKOLKICENKSGCEYCKAKDVSTORLTG 1980
 QY 1981 NSONMNASIDDEKFEYEGKNCQVPRGPVPRVREPSPRSLISKADASKKAKTAPPR 2040
 DB 1981 NSONMNASIDDEKFEYEGKNCQVPRGPVPRVREPSPRSLISKADASKKAKTAPPR 2040
 QY 2041 QPKKVENLTTEMAQOTRRRAAQOTRRRTSTATTTESDVGTWAKALISNKPDSRGIEGC 2100
 DB 2041 QPKKVENLTTEMAQOTRRRAAQOTRRRTSTATTTESDVGTWAKALISNKPDSRGIEGC 2100
 QY 2101 NPKTYGYPRWGCYVSKSEBENGICMPPRKKLCINNIQYLYNETENKRDNDIKEAFIK 2160
 DB 2101 NPKTYGYPRWGCYVSKSEBENGICMPPRKKLCINNIQYLYNETENKRDNDIKEAFIK 2160
 QY 2161 CAIEFOFLMKYLIENPAENELONGTIDDEKRIIMYTYGYKMPFGTDSINCKIT 2220
 DB 2161 CAIEFOFLMKYLIENPAENELONGTIDDEKRIIMYTYGYKMPFGTDSINCKIT 2220
 QY 2221 TYTNSVTTILNENKKKODKKODELRKIFWEKKNKFIEGMIYGLTYHLTDENEKIKR 2280
 DB 2221 TYTNSVTTILNENKKKODKKODELRKIFWEKKNKFIEGMIYGLTYHLTDENEKIKR 2280
 QY 2281 DMYQYDMTKLIPSLSEFYKRPOLFWETEMABEFCNKKREOLKLEBAGCKEYECNSND 2340
 DB 2281 DMYQYDMTKLIPSLSEFYKRPOLFWETEMABEFCNKKREOLKLEBAGCKEYECNSND 2340
 QY 2341 GKTQCEABACVYQNFIKKMKTEYERQREKFKDKGKRYKDYPSERDIEKATCAHEYL 2400
 DB 2341 GKTQCEABACVYQNFIKKMKTEYERQREKFKDKGKRYKDYPSERDIEKATCAHEYL 2400
 QY 2401 NMKLELCGNKDCSCMQKPSOLPKTTQOSSDANDMPESLDYVEEFNKCECPBELSKR 2460
 DB 2401 NMKLELCGNKDCSCMQKPSOLPKTTQOSSDANDMPESLDYVEEFNKCECPBELSKR 2460
 QY 2461 GSMIHTKKTITEKIPMNCVEKAAAYISKAEENMDITLKEKPIPIESTKEKESKNMTNN 2520
 DB 2461 GSMIHTKKTITEKIPMNCVEKAAAYISKAEENMDITLKEKPIPIESTKEKESKNMTNN 2520
 QY 2521 NPODPKPYAPDKYIGRRNRCENRENRKVDYEMKCYNSKRYOKKKRVCPRREHMC 2580
 DB 2521 NPODPKPYAPDKYIGRRNRCENRENRKVDYEMKCYNSKRYOKKKRVCPRREHMC 2580
 QY 2581 LRLNDEIKTERLKDSNYLLKMYRTARNEGIDIIKFNSENGCAMPICDTMKYSPADLG 2640
 DB 2581 LRLNDEIKTERLKDSNYLLKMYRTARNEGIDIIKFNSENGCAMPICDTMKYSPADLG 2640
 QY 2641 DIVRGDMRLRGYLPVPEIKLYVEEYIYGKRNKNGKANKYNDVQTRSAWMDANKRD 2700
 DB 2641 DIVRGDMRLRGYLPVPEIKLYVEEYIYGKRNKNGKANKYNDVQTRSAWMDANKRD 2700
 QY 2701 IMKAMTCAPEDAKLFRRKRMGCFERITLIDCKGKNDPPVDYIIPORRMWTENSEY 2760
 DB 2701 IMKAMTCAPEDAKLFRRKRMGCFERITLIDCKGKNDPPVDYIIPORRMWTENSEY 2760
 QY 2761 CKAALMELEKFKKSCDHCTSDCKNDYDENKCEQCKTRCOEYKKNVLLKWSLFDIOSNK 2820
 DB 2761 CKAALMELEKFKKSCDHCTSDCKNDYDENKCEQCKTRCOEYKKNVLLKWSLFDIOSNK 2820
 QY 2821 YKELYEOPITYTKISTYDHQNFQVKLTKFKSECSVESFSYLETSKCLNYKKNENDGSS 2880
 DB 2821 YKELYEOPITYTKISTYDHQNFQVKLTKFKSECSVESFSYLETSKCLNYKKNENDGSS 2880
 QY 2881 NIKTYAFEEETPKSYKACSCITLPSKNPLDNCPTDQNKDGCKELQOTFCSKNDYDNLLND 2940
 DB 2881 NIKTYAFEEETPKSYKACSCITLPSKNPLDNCPTDQNKDGCKELQOTFCSKNDYDNLLND 2940
 QY 2941 WNAAYVLNSSDNDKGVLLPERRRHLCITRTAYNYRKGKELIKKLLTSAFSQOGLLQ 3000
 DB 2941 WNAAYVLNSSDNDKGVLLPERRRHLCITRTAYNYRKGKELIKKLLTSAFSQOGLLQ 3000

QY 3001 KYKSEELCEFAKRYSAOYSDIILKGTDMMDTSLSEKIKKIFETSSEATENKRTWENNR 3060
 DB 3001 KYKSEELCEFAKRYSAOYSDIILKGTDMMDTSLSEKIKKIFETSSEATENKRTWENNR 3060
 QY 3061 ROIWHMLGCKYKATSVTLDEGMCOLPKDEFNOPRLIEMAKOACREKKAHVSOLKT 3120
 DB 3061 ROIWHMLGCKYKATSVTLDEGMCOLPKDEFNOPRLIEMAKOACREKKAHVSOLKT 3120
 QY 3121 KCPRSNEDNEASELLROPQCONDIRKYISLNTILIKNTMENLIRYKOLKODQSSGNIDNK 3180
 DB 3121 KCPRSNEDNEASELLROPQCONDIRKYISLNTILIKNTMENLIRYKOLKODQSSGNIDNK 3180
 QY 3181 PSEENVOYIKSKDSOCALENDINEIYVGTCKNNENNEFEVILKLYPGLYFEDETHKN 3240
 DB 3181 PSEENVOYIKSKDSOCALENDINEIYVGTCKNNENNEFEVILKLYPGLYFEDETHKN 3240
 QY 3241 HVLDSNKEEEOYVRKALYFETPHVDSFOAQLFSTHRVAQVDPKNDILKSSISVIVS 3300
 DB 3241 HVLDSNKEEEOYVRKALYFETPHVDSFOAQLFSTHRVAQVDPKNDILKSSISVIVS 3300
 QY 3301 ALGLIALHFMKKKFKSSVDLRLINIPQGEYGMPTLESKNRYIPYSGPYKGTIYMEG 3360
 DB 3301 ALGLIALHFMKKKFKSSVDLRLINIPQGEYGMPTLESKNRYIPYSGPYKGTIYMEG 3360
 QY 3361 DTSGDEDKYMDLSSSDITSSSEYEBELINDIYVPSGPKYKLEIYVLEPSKRDIPSD 3420
 DB 3361 DTSGDEDKYMDLSSSDITSSSEYEBELINDIYVPSGPKYKLEIYVLEPSKRDIPSD 3420
 QY 3421 TPSENDTPRTNRFIDDEENELKDFVSOYLPTEPNNNYKASDIIPMTEPWTLYSDNEEK 3480
 DB 3421 TPSENDTPRTNRFIDDEENELKDFVSOYLPTEPNNNYKASDIIPMTEPWTLYSDNEEK 3480
 QY 3481 PFIIISHRDLYTGKEISYNNINNSTNTNDIPNNARNDYRGIDLINDSLVNLILIYMM 3540
 DB 3481 PFIIISHRDLYTGKEISYNNINNSTNTNDIPNNARNDYRGIDLINDSLVNLILIYMM 3540
 QY 3541 KY 3542
 DB 3541 KY 3542

RESULT 2
 AAW22475
 ID AAW22475 standard; Protein; 3060 AA.
 XX
 AC AAW22475;
 XX
 DT 12-SEP-1997 (first entry)
 XX
 DE Plasmodium var-7.
 XX
 KW DBL gene family; SABB; static acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 XX
 OS Plasmodium vivax.
 OS Plasmodium falciparum.
 PN W09640766-A2.
 XX
 PD 19-DEC-1996.
 XX
 PF 07-JUN-1996; 96WO-US09508.
 XX
 PR 07-JUN-1995; 95US-0487826.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 XX
 DR WPI; 1997-052231/05.

DR N-PSDB; AAT72882.

PT New malaria vaccines - contains cysteine-rich DBL family protein
binding domains homologous domains of the Duffy and sialic acid
binding proteins

PS Claim 8; Page 61-67; 96pp; English.

CC This sequence represents var-7 of Plasmodium. Var-7 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and sialic acid binding protein
CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
CC var family of genes modulate erythrocyte invasion and antigenic variation. Of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and
CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the
CC invention. The compositions are for the treatment and prevention of
CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
CC family of genes having homology with conserved regions of DABP and SABP.
CC The compositions are used for the treatment and prevention of malaria.
CC They are also used in the preparation of vaccines for inducing a
CC protective immune response in a mammal to Plasmodium merozoites
CC (especially Plasmodium falciparum or Plasmodium vivax).

CC Sequence 3060 AA;

Query Match 18.0%; Score 3489; DB 18; Length 3060;
Best Local Similarity 26.7%; Pred. No. 2.5e-200;
Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

QY 9 IIKMGAAASLEGDASPIIKSHKASARVLEKRNIRHPSK-YAKEHVDISIKGDLTKA 67
DB 4 LAKMGKREAGSDDIED-----ESAKHMFIRIGKDYDKYKBAKRGKGLQRLSPA 56
QY 68 EFRGSPSTPVNKHNYYPVPCMLDKHKNLRYDDVNLHPHGHEQRNFDESESEC-G 126
DB 57 KEKNESDPQTPED-----PCDLIKHYHNV---TTNVLNPCADSDVFRSEYSGCQTH 108
QY 127 NKIRYKRRKNDIAICAPRRHRMCKNLALNDINTONTIHLLGNVLTAKYEGESIVNN 186
DB 109 NKIKSOGQDNKGACAPRYRLHVCDQNLQEPRIKITHTHNLVYCAAKFEQSTIOD 168
QY 187 HP-HKGT-----SDACTLARSEADIGIVRGIDMFKRNVD-----KVEGLREVEFKI 235
DB 169 YPKYATIGDSQICTMLANSFADIGIVRGIDLYGNPQIKQRQDLNNLKTIFGKI 228
QY 236 HD---GMEDEKNDYNDPGSGNYTKLREAWNVNKNKWEAATCDASYKSGYFQMOSENT 292
DB 229 YEKLGAGARVGN--PE-----FFKLRBDWMTANRETYWKAITCNA-WGNITYF-HATNR 280
QY 293 PLFSNPKGCHGKVPYMLDVPQYLRMFDEWGEFCKRNIRIKKYVDSGR---NDKE 348
DB 281 GERIKGYRCNDVDVYFDVYPOYLRFWEEMAEDEFCRKKKKIKDYVRNCRGKDCKED 340
QY 349 RLYCSHNGHDCITTIWKGILHLNKCCTGCTKCKVFWMLGNOGEAFKKQKKEVEKEIQ 408
DB 341 R-YCSRMGYDEDEKTRAKGLRYGKQCSCLYACNPYDWMNNOKEQDQKKKKYDEIRK 399
QY 409 SY-----LSNDNKFVNINSE-YKQFYEKLKETQVATNTPTNLINEGKYC--- 454
DB 400 KYENGASGSGRQKRDAGTITTYNDGYEKRYDELNSERYTVKFLKLSNEITCIRKV 459
QY 455 --KGLGEGKQITFTNSA-----DDKGLFYRSEYQVCPDGCYKDGKILKYTH-KSDND 504
DB 460 DEEGGIDTFKVNDSSTGASGTVNESQGFYRKYQCPYCGV---KVNNGGSSNE 515
QY 505 RERVNN-----EDYKPPWGVKPTNITVLYSGNEQDITQKLEKFCN-----SS 547
DB 516 WEKNNKGCKSKGLYEPKPKEGTITITLISGKHCHDIEKLNKFKCEKKNQDITINSGGSG 575

QY 548 TNYKDKNN-----QKMECYKDENINCKLEQNTLEINNDPK-----II----- 586
DB 576 TGGSGGSGNSROBELYEEMKC-YKGEDVYKVGHDEDEDEEDYENVKNAAGCLIAKNNKKE 634
QY 587 -----SEHNFELWYTLIRDTIKKNDKIKTQI-NNTTTHC-IDECNRNC 629
DB 635 ECGNTSEKPEDELQKTFNFFYYWAHMLKDSIHKKKILQRLQNGNRIKCGNNKNCND 694
QY 630 LCFDRVYKQKEEWMNSIKKLFTRKNQI-----QSYSNINLFEGEYF--- 672
DB 695 ECFKMITQKKEWCKIVQHF-KTQINRGSGSDMTALIPRDHYLYQVNLQEFELKGD 753
QY 673 -----FKYMDKLDEDAKKEMLNIRKKKEFNLENRD-----YLENAITELL 718
DB 754 SEDASEESENLDABA-----ELHNLRIETSEEDNNOEASVGGVTEQKINMDKL 807
QY 719 DHAKEATATC-----KDNNTNACETSHNATNPGVPRPGTQ-PTKNIKELAQ 766
DB 808 NYEKDEADICLETIHDEDEBEKKGDNCEIDEGENFRYNPCSGESGKNRYVLANKAYQ 867
QY 767 YKRSAYEARRGKLAHLKKAHESITYRGGRRDEKDNLCRIMIKHSNRLGFSNGPCD 826
DB 868 MHHKAKTQLASRAGSALRGDISLAQFNKNGSYLQICKINENYSNDSRGSGPCT 927
QY 827 GKGTGDIQTRPVVGTMEVDEPHMKRDHEDVIMPPRRRHICTENLHLOTDHPLANGNT 886
DB 928 GK-DGDHGVYMRKISTEENSIEGKKQTSYKNVFLPREHREKSTNLELDV-----GSV 980
QY 887 V-DLVNNSFLGDVLLSAKYBANKIIRAYKERNKMLKPKETVTD---KHOTICRAIYS 942
DB 981 TKNKASHSLGDVOLAAKTDAEILIKRYKQNNI-----QUTDPIQKQDEAMCRAIYS 1036
QY 943 FADIGDIIRGRDIMERNDAMKLOGHLEYVGNTHKSLKG-KGNDKYND---APKYLK 998
DB 1037 FADIGDIIRGRDIMERNDAMKLOGHLEYVGNTHKSLKG-KGNDKYND---APKYLK 1096
QY 999 RENNWEANRAKVPAMKCDIKYLLKDKSGHSTQSGYSQSDHPPLDIPQIKRWMTEMA 1058
DB 1097 RADWENRHRQVWAMK-----ATKILIC---PGHVPDYIYQRLRMWTEMA 1141
QY 1059 EWCCKYQKKEVDKLEKCEKCKDKONGQCTKESGTCCTCTACENYNDIIGLMEOWN 1118
DB 1142 EWCCKYQKKEVDKLEKCEKCKDKONGQCTKESGTCCTCTACENYNDIIGLMEOWN 1199
QY 1119 IISDYKELHEQAOQSVNSNGIEASSYAKNHD---RNVLEFISELYQOQNGK---SK 1171
DB 1200 KISDKYMLVLAQKTTSTNG-----RTVLGDDDPDYQOAWDFLPIHKASIAARVLVKA 1255
QY 1172 SGTSDSAVIGTNTTENVGAYLHDTGNFDDCOSONEFCDEK-----SDGKNEKYAFR 1225
DB 1256 AGSTELTAAAPITPYSTAGIYHOEIGYGQEOFOFCEKKGANSTSTENKNEYTRK 1315
QY 1226 DKPODHGACGCKSGSKPTVQIKTKKKAERKTECKTYVNDIILEKNDGKQVEDCPKRN 1285
DB 1316 QPPEVATACDCLNRSGT---EPKKKEENYVSAKIVIEKLIGNKGRITTYGECBPXES 1371
QY 1286 SNGIPDMQC-GNINLYEDPRVCAPPRRQKLYAFHLANDNEIKKLOSVMLKEAFITSA 1344
DB 1372 ---YPMDCKNNDIDISHD-GACMPRRQKLCIYLAHESQTEIKITDMLKDAFIKTA 1427
QY 1345 ETEFSWYTKSK-DGAGCNELDKELGKILPPAFLRSMFYTFGDRDLFGTDISKGHG 1403
DB 1428 ETLISQWYKSKNDSEAKTIDR---GLIPSOFLRSMYTFGDRIDICLNTDISKONDY 1483
QY 1404 SKLEQIDSLFKRGDQKSPNGKTRQEMWTEHHEHETMEALCALVYIGAKKD---FTERY 1460
DB 1484 AKAKDKIGKFFSKDSGSKSGSLRQEMWTKTNPGRIMKGMCAITTYVTDTRRRIRKIDY 1543
QY 1461 GYNNVFFSK-STTLEEFKAPRQFLWLTWYDDCYTRQKLIKDVQEKCS-NDQLKCD 1518
DB 1544 SYDKVNOSONGNPSLEEFPAKPOFLMMIEMGECECAERQKKEENIKKACNENINSTOQN 1603

QY 1519 T---EONKCEDYVYKMK-KKKEMIPQDKYKDEBERKKRPFDRKHIGVMTDTYTGNATDY 1574
 Db 1604 DAKHRCNOACRAOYOEVENKKEFSGOTNNFVKANVOPODDEYKCYEKD----- 1654
 QY 1575 LNRKFTASGDRKGSASVYORNIOLLEKQAYYDADHCGCTKFIENDDKTYTINSSKDKK 1634
 Db 1655 -----GVOD-----IUGNETYLOK-----CDNMK-CSC----- 1676
 QY 1635 GLVKEANTGAIKQNKGNPNNNYNNLKEITBEDVLEPPSRRLRICEHALDGNVYDPEVKDENG 1694
 Db 1677 ----- 1689
 QY 1695 KRRLMEVATTEGYNLQYKREKKEKIKITSDAHKISTEYVPCSSANKYSFYDLRDLITLIG 1754
 Db 1690 -----PEGKY-----AHKYP----- 1699
 QY 1755 DNLEDEKOTEBLKKIFNNKNGTVGKGSJTGNGSTARKPFEMENKECVANNMICGY 1814
 Db 1700 ----- 1699
 QY 1815 KRGRODNGSNGSARSDDELKKGCVSPDDYPMKRNDEGTAYOPLRMFAEMGEDECKHR 1874
 Db 1700 ----- 1699
 QY 1875 EKELEKLVGACNDYTGDMEDKRRKCTDACTOYKRFISEMKPOYKQIKKYGKNNKIYS 1934
 Db 1700 -----EKC----- 1702
 QY 1935 EHFVADAEADAREYLDKOLKICENKSGDCEYKCMKDVSTQRLTDGNSQMPASLDEPK 1994
 Db 1703 -----DC----- 1704
 QY 1995 EYEGKCNCOVPRGPPRYRRETPSPRYSLISKATVASKREAKTAPPTQPKVENLITTEMR 2054
 Db 1705 -YOGK---HVPSTIP-----PPPYQPO----- 1723
 QY 2055 QRTIRAAOQTRKRTSTATTESDVTWYKAILSNKPDSSRGIEGNCNPTTYQY---PKWG 2112
 Db 1724 -----PEALPYVYVCSIVKTYLTK-----DTNNEFSDACGLK-YGKTAPSSWK 1764
 QY 2113 CI-----VGSKENENGICMPRRKKLICINNIOYLANET-----ENKRD 2151
 Db 1765 CIPSDRKSAGANTGKSSDGSICIPRRRRRLYVGKLO---EWATLPLQEGAPASHSRA 1822
 QY 2152 NDIKEAFICAIETOPMLKTYIE-NPAENE-----LONGT 2188
 Db 1823 DDLRNAPFIOSAIETEFMLDRYKEEKKPOGDSQALSOULTSTYSDEEDPDKLLONGK 1882
 QY 2189 IPDEFKRYMYTYGDKDMFEFGDISDKKIIVTNSVTTILNENKKKKODKKDEELK 2248
 Db 1883 IPDEFKRYMYTYGDKDMFEFGDISDKKIIVTNSVTTILNENKKKKODKKDEELK 1942
 QY 2249 I-----FWENKKEFTWEGMIYGLTY-----HLTDNEKKEKIR 2280
 Db 1943 ILPKNGSTPLVPKSSAOTPDKMMNHEAESIMKMICALVYTEKNPTDSARGDNKIEKD 2002
 QY 2281 DMT-----QYN-DMTKL-----TPSLEEEYKRP 2302
 Db 2003 EYVEKEFFGSTADKHGTASTPTGTGYKTOYDEYKVKLEDTSGAKTPSPASDPLILSDVLRP 2062
 QY 2303 OQLRMFTMAEEFCNRKREBOLKLEAGCKEYE-----CN---GSN 2339
 Db 2063 PYRYLIEBMOONCKRRKHLKLOIKHECKYEEBGGSRGCGITROYSGDCEACONEMLPKN 2122
 QY 2340 DGTQOE---CAEACVYQNFITKKWTEYEROREKFKDKD---GKKYKDYPTSTERD 2390
 Db 2123 DGTVPDLKPSKCAKPCSSYRKWIESKGEPEKEKAYEQDKCVCVGSNNKD----- 2174
 QY 2391 EKATCAHEYILNMLKELCGKDCSCMOKPSSQLPKTTQOSSOSDANDMPSLDYVEBEFN 2450
 Db 2175 -----NGFCETL-----TTSKADDFLTKLIG----- 2195
 QY 2451 KCEBPELSKSGSMI-----HTKRTIEP---KIPMNCVKAAYILSKAEANMDITLKE 2500

Db 2196 --PKRPNNVEGKITFDDDKTFKHTKD-CDCPLKFSVNC----- 2230
 QY 2501 KFIPIETKEKESKNSGTNNNNPCDPKPKPYAPDKYIGRRBPCENREBRKVIDYEMKCYKN 2560
 Db 2231 -----KKDECDNS-----KJDCRN-----KN 2247
 QY 2561 SKFEQKKRVCVPPRREHMLCRLNDEIKIERLDSNYLLKMYRTARNEGIDIKNNFS- 2619
 Db 2248 S-----IDATDIENGVSIVLEMRVADS-----GFND 2278
 QY 2620 --ENGCAAMPICDTMYKTSRADLGIYRGTDMLRIGITLPEVEIKLYKVEYIGKRRNN 2677
 Db 2279 GLENAC-----RGAG-----IFEGI----- 2293
 QY 2678 KGRNKYNDVOTFRSAMWANDANRDKIMKA-----MTCKAPEDAKLFRKGRMDGERITLIG 2731
 Db 2294 -----RDEMKCGRVCGYVVK-PENVNGEAKG----- 2321
 QY 2732 DKCGHKDDPPVDYIPQFRFMMTSEYCKALMELEKFKKSDCHTSRCKNDYDEN 2791
 Db 2322 -----HIIQIALVYKRWVEYF---FEDYNNIKH---KISHRIK---G 2355
 QY 2792 KCEQCKTRQOEYKNEFVLLKKSLEFDIOSNKYKELYEQPIYTKISTYDHOVQKLTFS 2851
 Db 2356 EISPC-----INCYEKW---VDOKRKEKETER-----FKD 2385
 QY 2852 ECSVESFSEIETHETSKCLNKFENENDSSNIRTYAEETPK-----SYKEA 2897
 Db 2386 Q-----YK-NDNSDDNVRSEFLETLIPQITDANAKKNVILSKFSGNS 2426
 QY 2898 CSCTLPKSNPLDNCPTDOKNDGCKELQTFCSKNDYDNNLDMNAYLVLNSSDNNKGYL 2957
 Db 2427 CGCSASA-----NEQNNK----- 2440
 QY 2958 IPPRRHLCRTPTATVAYNRKGDKEILLKLLTSAFSOGOLGKYSEBELCEAKKYS 3017
 Db 2441 ----- 2440
 QY 3018 ADYSIILKGTDMNDTSLSEKIKKIFETSNEATENRKRTWMENNRQIMHAMLCYKIATSK 3077
 Db 2441 -EKKAL-DCMLKRLKDKI----- 2457
 QY 3078 VILDEMCOLPKDEETNOFLRLILMAKQACEKKHVSLSKTKC-----PRSNEDNEFA 3132
 Db 2458 -----GEC-----EKKHOTS-DRECSDTPOPTLEB----- 2483
 QY 3133 SELLRQPGCONDIRKYSILNLIKNTMENTINIKYKLOKQOSSGNINDKPSSEENVQSYKS 3192
 Db 2484 -ETL-----DDDIETEBAKNNMPKICEVY-LKTAQOODE-GGCVPANSEBPA-----A 2530
 QY 3193 KDSQALDELNDINEIYVTRKNNENNEFEVKLYLQGLYFVEDETHKNHVLNDINKEEQ 3252
 Db 2531 TDS-----GKETPEQTVLK-----PEEA 2550
 QY 3253 TVRPKALYFTPHVDSFYQAPLESTHRVADYKN-----DLKSSISVYIYALG 3303
 Db 2551 VPEBP-----PPPOKEAPAP-----IPQPPPTPTQLDNPVHVALVATVSTLMSVG 2599
 QY 3304 L-----IALHEMKKFKSSV-DLRLIINIPQEGEMPLESKNRYIPRSGPYGKTYIYM 3358
 Db 2600 IGFATYTFYFLKKTSSVGNLFOIOLIPKSDYDIPKLSPNRYIPTSTSKYKGRITYL 2659
 QY 3359 EGDY-----SGDEKRYMMDLSSDITSESEYEELINDIYVPGSPKYYTLLEVLLEPSKRD 3415
 Db 2660 EGDSDGSDGYTDHY-----SDITSSSEYEEMDINDIYVPGSPKYYTLLEVLLEPSGN 2713
 QY 3416 IPSD-----DTPS-----NTPRNRKIDENMLKIDYVSYOL--- 3449
 Db 2714 TTASGNNTTASGNNTTASGNNTPSDQNDIIONGIGIBSKITDENMQLKDEFISOYLQSE 2773
 QY 3450 PNTPEPNNTKSAIDIPNTEPNTLYSDNPEEKPLISIHBDLTGKEISY----- 3499

QY 1119 IISDKTELHEQOMSVNSGIEASSTAKNHD---RNVIEITSELYOONGK-----SNK 1171
 DB 1200 KISDKTNLYLQAKTTSTNG-----RTVLGDDDDPYQOAWDELTPPIHKASIAARVLVKRA 1255
 QY 1172 SGRSDSAVIGTWTTEYNGAYLHDGNDPDOSONEFODEK-----SDGKNEKYAPR 1225
 DB 1256 AGSPTEIAAAPTTPYSTAGYTHOEIGGCOEOTOFCEKKHGANSTSTTKNKEKTEFR 1315
 QY 1226 DRPODHDGACGCKSGSRPTFVQJTKRKAEEKTECVNDILKENDGKQVEDCHPKKN 1285
 DB 1316 QPEPEVATACDCINRSOTE-----EPKKKEENVESACKIYEKILGKNGRTVECAPKES 1371
 QY 1286 SNGYPRMOC-GININLYEDPVPCPRPRKQICVFLANDNEIKKLOSOVNLKEAFITSA 1344
 DB 1372 ---YPPMDCKNNDISHD-GACPPRRKQICLYIAHSEOTENIKDDMDKAFITAA 1427
 QY 1345 ETEFSWYYSK-DGEGNELDKELKEKIPAPFLRSMEFYEGDYRDLFTGDISKGE 1403
 DB 1428 ETELSMOYKSKNDSEAKILDR-----GLIPSOFLRSMMTFGYRDLCLNTDISKQNDY 1483
 QY 1404 SKLEQIDSLFKNGDQSPNGKTRQEMWTEHSHIWEAMLCALVIGAKKD-----FTENT 1460
 DB 1484 AKAKDKIGKFEKSKDGSKSPGSLRQEMWKTNGPEIKMGMLCALTKYVTDITDNKRKINDY 1543
 QY 1461 GYNNVFSOK-STTLEBFAKRPOLFLMTEWYDYCYTRQKYLKDOEKCS-NDOLKCD 1518
 DB 1544 SYDRVNOSONGNSPLEEFAKPOFLRMLEMGEFECAEROKKENIJKDACNEINSTOQCN 1603
 QY 1519 T---ECKKCEDYVYKWK-KKEMIPQDKYKDERDKKRDROHIGVMTDYGTATDY 1574
 DB 1604 DAHRCNOACRAVOEYENKKEFSQGTNNFVLKANYOPDPEYKGEYND-----1554
 QY 1575 LNRKFTASCGDKPSASVVOIRIOLLEKQAYADAKHCGCTKEIENDKYTNISKDKCK 1634
 DB 1655 ---GVOP-----IGNEYLLOK---CDNNK-CSC-----1676
 QY 1635 GLYKEANTGAIKQONKGNPNYNNLKELEBYLPPSRRLKICFALDQNTYDPEYKDENGL 1694
 DB 1677 ---MOGNVLSVSPK---1689
 QY 1695 RKRLMEVAATEGYNLGOYKYEKKERIKITSDAHKSIEYEPSCAMKSYFDRLDILGI 1754
 DB 1690 ---PFGKY-----AHKYP-----1699
 QY 1755 DNEDEKQTEENLKKIFENKGTSGVKGSDSTGNGPSTARKEFWNENKECVNNAITGY 1814
 DB 1700 ---1699
 QY 1815 KRGRDQNSGNSARSDDELKCKGSPSDDYPMGKNRDEBTAYQFLWFAWMBEDPCKHK 1874
 DB 1700 ---1699
 QY 1875 EKELEKLVGACNDYTCGDNEDKRRKCTDCTQYKFISEWKPOYEQIKKRYGENKDIYS 1934
 DB 1700 ---EKC-----1702
 QY 1935 EHPVAKDAEDAREYLDKOLKIKCENKSGDEYKCMKDVSTQRLDNGSNQNPASLDEPK 1994
 DB 1703 ---DC-----1704
 QY 1995 EYEGKNCQYPRGPRPRRRETPSPRVLISKATASKKEATAPPTKOPKVENLITTEMA 2054
 DB 1705 -YOG---HVPSTIP-----PEPPVOPO-----1723
 QY 2055 QTRRRAAQTRKRTSTATTESDVTGTMKAILSNKPDSPRGIGCNPRTYGOY--PKWG 2112
 DB 1724 ---PPAPTYVDVCSIVKTLK---DTNNFSPACGLK-YGKATAPSSWK 1764
 QY 2113 CI-----VGSKENENGIKMPPRRRKILCINNIOYLANET-----ENKRD 2151
 DB 1765 CIPSDTKSGAGATTKGSGSGSICIPRRRRRLYVGLQ--EWATLALPOGGAAPSHSRA 1822
 QY 2152 NDIKFAITKCAITQFLMLKYTEE-NPAEN-----LONGT 2188

DB 1823 DDLRNFIOAALTEPFLDRKKEKPPGDSQOALSOLTSTYSDEEDPPDKLLQONK 1882
 QY 2189 IPDEFRIWYTYGDKMEFFGDISNKKILVTNSVTTILNENKRRKODKDELEK 2248
 DB 1883 IPPDFRLMFTYGLDRDLIVHOGNTSDSGNTNGSNNNIVLEASGNKEDMKIOKIQ 1942
 QY 2249 I-----FWEKRRKFLWEGMIGLYY-----HLTENEKIR 2280
 DB 1943 ILPKNGTPLVPSASQTPDKMNEHAESIMKMICALYATEKNPDTARGDENKTEKD 2002
 QY 2281 DNY-----QYN-DMTKL-----TPSLSEFVRK 2302
 DB 2003 EYIEKFFGTAHNGTASPTGTGYTKQYDEKVLKEDTSGAKTPASDPLPLSDVLRP 2062
 QY 2303 QFLMFTMAEFCNRRKREKOLLKLEAGCKEY-----CN---GSN 2339
 DB 2063 PYFRVLEMGONCKRRKRLKQAIKHECAYEENGSGSRGIGITROYSGGECNEMLPKN 2122
 QY 2340 DGTQOE---CAEACVYONFIKMKTEYEROREKFKDKD---GKKYKDYPTERTD 2390
 DB 2123 DGTVPDLKPSCAKPCSSYKWTIESKGEKEKAYEQKDCVGSNKH-----2174
 QY 2391 EKATCAHEYLMLKELCGKDCSCMQPSQPKTTOOSQSDANDMESPILDYVEERN 2450
 DB 2175 ---NGFCETL-----TTSKARDPLKTLG-----2195
 QY 2451 KCEPELSKKGSMI-----HTKKITEP--KIPMNCVKAAYLSKEANNITLKE 2500
 DB 2196 -PCKNNVNGKITIPDDKTFKHTKD-CDPCLKFSVNC-----2230
 QY 2501 KFIPIESTKESKNSWNTNPNPCDPKPYAPDKYIGRRNCENRENERKVDYEMKCYAN 2560
 DB 2231 ---KKECDNS-----KGTDCRN-----KN 2247
 QY 2561 SKFYOEKKRKCVRPRRHHMLRLDEIKIERLKDQSVTLKVMRTARNEGIDILKNPNS- 2619
 DB 2248 S-----IDATDLENGVDSVLEKRSVADSRS-----GFNGD 2278
 QY 2620 -ENGAMNPICTMYKSFADLDIYAGTDMRLIGLYPEVEIKLYVEYIYKWRNNK 2677
 DB 2279 GLENAC-----RGAG-----IFGI-----2293
 QY 2678 KGRNKYNDVOTFRSAMWANKRDIWA-----MTCKAPDADLFRKGRMDGEFRLTIO 2731
 DB 2294 ---RKDEMCRNVCYGVVCK-PENVVGEAKG-----2321
 QY 2732 DKCGHKDDPVVDYIPORFAMTMSSEYCYKALMELEKFKSCDCKQTSDRCKNDYDN 2791
 DB 2322 ---HIIQIRALVRYWEYF---FEDINKIKH---KISHRIKN---G 2355
 QY 2792 KCEQKTRCOEYKNTVLMKSLFDIOSNKYKELYEQPIYTKISTYDHVONFVOKLTFKS 2851
 DB 2356 EISPC-----INCVEMK---VDOKRKEKKEITER-----FKD 2385
 QY 2852 ECYVESFSEYLTHTSKCLNTKFNENDGSSNIRIYAEETPK-----SYKEA 2897
 DB 2386 Q-----YK-NDNSDDDVRSFLETLIPQIDANAKNVIKLSFGNS 2426
 QY 2898 CSCTLPKPNLNDOPNDONCKELOTFFPSKNDYDNMLDMNNAVVLVNSDDNKGVL 2957
 DB 2427 CGCSASA-----NQNKNNG-----2440
 QY 2958 IPPRRRLCTRPIYANYRRGDKELIKKLLTSAFSOGOLLGOKYKSEBELCEAMKYSY 3017
 DB 2441 ---2440
 QY 3018 ADVSDIITKGMNDTSLSEKIKIIFETSNBATERKRTWENNROITWAMLCGKIATFSK 3077
 DB 2441 -EYKDAI--DCMKIKLKDKI-----2457
 QY 3078 VTIIDEGWOLPQDEBTQFLRWLLEMAKQACEKKHYSDSLKTRC-----PRSNEDNFEA 3132

Db 525 SKLELEKTFKCDGKIK---NDIKCHIDDNG-----TDDOTDSDNDVLDGMCULTE 576
 QY 583 PRIISFHNFEELVWYLLADTJIKWMDKLCTCIN-NTTTCIDECNRCCLCFRHWKQKEE 641
 Db 577 DKIMSYNAFMMVHMDLDSIKWDEHRCJINKDKGKICJGCKNKKJCFOKWMVQKKT 636
 QY 642 EWSIKKLTKKKNNQOYSYNNINLEFEGYFRVMDKLKDEAKKELMENIKRKNERS 701
 Db 637 EWGKIKDHRKODIPKDM-----KDKK----- 655
 QY 702 NLENNRDYENAI--ELLDLKETATICKDNNTNEACETSHNATNPCVPRGQTPTK 759
 Db 656 ---THDDFLQTLMKDLLEIITDYT---GDANE----- 683
 QY 760 NIKELAQYKRSAYEBARNRGLKHLKGAHEGILYKRGRRKDKDNLCKIMIKSHRNIG 819
 Db 684 -IKRI-----EALLBQA---GVGGIDFALAGITGFVAE---KDTIDKLLQHEBK--- 729
 QY 820 FSNGPCDCKGTGDIGOTRFRVGTMEVDEPHARKHEDYIMPPRRRHICTSNLEHLQTD 879
 Db 730 ---FADKCLKTHTDTCPPQ-----ED 748
 QY 880 HPLNGNIVDDLVNNSFLGVDLAKYEANKIIRMYEKENNLGPKREYDCKHOTICRAI 939
 Db 749 RSV-----ANSESATV-----PSPPADPK----- 767
 QY 940 RVSFADIGDILNGRDLMEHNGDMVKLOHLETVFGNIRKSLGKGNKNDAPKYLKLR 999
 Db 768 ---ATEEVDANASSDGEDDEE----- 787
 QY 1000 EKWMEANRAKWEAMKCDIKYLKDKSGHOSTOSSYCGSDHPLDIDYPOKLRMTMAE 1059
 Db 788 --- 787
 QY 1060 WYCKVOKREYDKLECKCKCKNDKNGOGCTK---ESGCTCKTEACNEYNDIIGLMKQ 1116
 Db 788 ---BEEDEBEEBAEVEYOEKTDSEATEVAAPSPPTQO--- 824
 QY 1117 WNIISDKYELHEQOAMSVNSGIEASTAKNHIDRNVLEPSELVQOONGSKNSKSTSD 1176
 Db 825 ---DGVPAQOEDDVYKCSI---VDKAL-----KGLD 851
 QY 1177 ESANVT---GTNTYENVAGAYLHDIGNPDOSQONFCECKSDGKNEKTAFAFDKQDDH 1232
 Db 852 DACTLKYGKGDNTTTEST---TKP----- 872
 QY 1233 GACGCKSGSKPTFRVOJKTKKKAEKDECKTYNDILKENDGKQVEDCHPKKNSNGYDPM 1292
 Db 873 GAAGTBSG---KDT----- 883
 QY 1293 OCGNINLVEDPVRCPMPRRKOKLCVHFL---ANDN----- 1323
 Db 884 ---GSI-----CVRPRRKLIVGKLDHMAGETTEAKSOETSOGQKTPSSGESSPSEK 933
 QY 1324 ---ETIKLOSQVNLKELAFIKSAAEFTFSWYTKSK-----D 1357
 Db 934 LPOGPPTETTKETPESSLHAFAVSPRLRFLRFLMHKFKEDQKMAQOAGATGOQTIICTLD 993
 QY 1358 GEGENEL-DKELKEGKIIPAFIRBMEYTPGDYRDLFG-NDI---SGHGBGSLKQJDS 1412
 Db 994 GGGEETPDLKLTGHLPPDLKROMFTYLDGYRIDLVGNTDIYVHTSGNEDMOIMEAIQK 1053
 QY 1413 ---LFKNODOKSPNGKT-----ROEWMTESHEIWEAMLCALV----- 1447
 Db 1054 KIRQIILPTSSSPSPRVQTOHSHVENPKRTWMNENGGKIMEGMCALYINNDTPSGTAP 1113
 QY 1448 ---KIGAKK-DEFTEN-----YGINNVKFS-----KSTILEEPKARPO 1482
 Db 1114 TOIOEVRTKIMDENSKNPKIPOYKYDOVKLDDTSKATGSPIPSEBKITPLDFTSRP 1173
 QY 1483 FLRLMEYDDYCYTROKYLKQVOEKC-----SNDQLKC-----DPECNKK---C 1525

Db 1174 YFRYLEEMGETPCKEKKKLLEKICECRDRTGHEHCSGDGDCITDADRDNDKFPDNLNC 1233
 QY 1526 EDYVYKMKKKEM--IPDQYKXDERDKKRFDRONIGVWVDTYGTNATYLNRRKFTASC 1583
 Db 1234 ROCHIOCRKRYRKMIDIKFDEYHROK-----KYOG-----EY----- 1265
 QY 1584 GPKPGASVYVQRNIQLEQOAYYDADKHGCGCTKFIENDKYNISSKDKGLYKAEANTG 1643
 Db 1266 -DK-----LTKDKSSGGDNNC-----CKDIEKHKSA 1291
 QY 1644 AIKMQKGPNNYNNLKELEDEVLFPSSRLRICEHALDGVNTPDEVENGKRLRMEVNA 1703
 Db 1292 VF-----LKEK-----HCNKGOTSEKKGQEDOL----- 1316
 QY 1704 TBGYNIGQYKKEKEKEKI-KTSDAKHYSYVPPCSAMKYSFYDLNDIILIGDNLDEKO 1762
 Db 1317 ---NKLDPRKIPQTFSPSTYKACQ-----YGV----- 1342
 QY 1763 KTEENLKTFKNGTSVSGSDS--TTGNPGSTARKFNNENKECYWNNMIC-----G 1813
 Db 1343 ---NCGNKRGRGNGCTTNNB-----PENKENDGGAASSTISILINDG 1383
 QY 1814 YKGRDGNNGSARSARDEDLKKGVSDDDYPM--GKNRDEGTAYOFLRWEFMEGDEDC 1871
 Db 1384 STNGATNGTGT--TDETLKEC---SDKTAFFGLKROEYTCOK-----KTVNGC 1429
 QY 1872 KHKKELEKLVGANDYTCGDNEDKRRKCTDACTQYKFISEWKPOYERQIKYGENKOR 1931
 Db 1430 ---NLTRVND--TYFDKO-----IVNEFFQRLARF---VHNYNLKHK 1467
 QY 1932 IYSEHVAADAEDAREYLDKOLKKEKNSGDCBYCCKMDVSTQRLTDGNSOMPSLSD 1991
 Db 1468 I---DPCIKKEKODKT---EHNCI-----NGCNI----- 1490
 QY 1992 EPREVEKNCQVPRGRPRRRETSPRVLISKATASKAEAKTAPTKQPKVENULTTE 2051
 Db 1491 ---KEC-----VR----- 1496
 QY 2052 MRAQTRRAAQOTRRRTSTATTESDVGTWVKAILLSNKRPSHGIEGCPNKTGYOYPRM 2111
 Db 1497 --- 1498
 QY 2112 GCIYVSKENENKICMPRRKILCINNIOYLANTEENKNDNDKEAFICAAIETQFLWL 2171
 Db 1499 LEIKG---NEWG---NIK---KHYNINSNDKET--- 1523
 QY 2172 KYIIEPAENELQNGTIDPEFRIMYYTYGYDKDM-PRGTDSNDKKITTYNSVTTL 2230
 Db 1524 ---IAYNKSYFVDDGLDITDYKKAQYV----- 1549
 QY 2231 NENNRKODKKDELEKIFWEKNKKFIWEGMAYGLJYHLTDE-NBEKIRIDNYOYNDMT 2289
 Db 1550 ---EDEKERRKIW-----GCTGH--DECSKKEBENK--NFTT 1580
 QY 2290 KLPSSLFEYVYKRPQFLRWETMAEFCNRKKEOLLKLEAGCKYEBGNSNDGTCQACA 2349
 Db 1581 NLISLQDKITSCQ-----NKH-----NPNKGT-----A 1604
 QY 2350 CVTYQNFKKMKTEYERQREKFKKDKKQYKYDPTSTENDIEKATGAHEVLMNKLKLELG 2409
 Db 1605 C----- 1605
 QY 2410 NKDCSCMQRPSSQLPKTQOOSDANDMPESLDYVPEEFNKECEPDELKSKGSMHTTKI 2469
 Db 1606 ---DPPSPPTBEETDPLDDPDPDPLDDOQHTBQPKFC----- 1640
 QY 2470 TEPKIRMKVEKAYYLSKAEENMDITLKEK-----FIPLESTK-EK 2511
 Db 1641 -PPPPMTCVEKIAKELRYAEAGKINNELKGNCKDENGRCNNYKKNGAVIGESCKPFO 1699
 QY 2512 ESKNSWTNNNPPCDPKFPAADKYIGRRNCPENENRPFKDYEMKCYKNSKFOE--KKRY 2570
 Db 1700 TYENSANNIN-----NKCKDNQNERFKIGKW-----NFKYIGTIRKDI 1738


```

QY 2571 CVPREHMCIRNDEIKERLKSNTYLLKVVBRARNEGIDITKFNSEGCAMPICD 2630
DB 1739 CIRPREHMCIRNDEIKERLKSNTYLLKVVBRARNEGIDITKFNSEGCAMPICD 1798
QY 2631 TMYKSFADIGDVRGDMIRIGYLPVEIKLYKVEFYIGKWRN-KN--KGRNKYNDV 2686
DB 1799 AMKYSFADIGDVRGDMIRIGYLPVEIKLYKVEFYIGKWRN-KN--KGRNKYNDV 1858
QY 2687 QTFRSAMDANKDITWKAMTCAPDADALFRGRND-----GFEITLLODKCGHKDDPP 2741
DB 1859 ---RSDMDANKHIMAMTCAPDADALFRGRND-----GFEITLLODKCGHKDDPP 1915
QY 2742 VDDYIPORFMRWTESEYECALMELEKFKSCDHC-KTSDRCNDYDENKCEQCKTRC 2800
DB 1916 DYDIPORFMRWTESEYECALMELEKFKSCDHC-KTSDRCNDYDENKCEQCKTRC 1975
QY 2801 QEKNFVLKWSLFDIOSNKYKELYEOPLY---TKISTYDHVQNFVKLTFKSEC-SV 2855
DB 1976 EYKKLHNMKLGFD---KYKEIYNE-IYNNKDSKINSNEYKFKLEKL---KDKCKEL 2027
QY 2856 ESEFYHETSKCLMKNFENNGSSNIRYAFETPKSKYKCEKSCGLTSKNPLDNCPIQD 2915
DB 2028 NSSDCIDEATICTYKYS-NSENKNHNNYAKKNPKYKCEKCKDAP--DPLDNCPRDS 2084
QY 2916 ---NKDGCKELQTFECSKNDYNNLNNMAYLYLNSSDNKGVLIPRRRHLCRPIYAY 2973
DB 2085 ATYERACNTLPTKLCESKTFNDDSWTSPVOTSPRNTGVLVPPRRRQCLAKNITT- 2143
QY 2974 NYRKGRD-EIILKKLLTSAFOSOGOLLGOKYKSEELCEAMKYSTADYSDITKGTDMDT 3032
DB 2144 KLRSEKIDDFAEMLTSAFOSOGOLLGOKYKSEELCEAMKYSTADYSDITKGTDMDT 2203
QY 3033 SUSEKIKKIF-----ETSENEATENKRWENNROIMHAMCGYIATSKVTLLEGMCQ 3086
DB 2204 APLDKLKTLYNLKAGDGTNEIKEDRGKWTENKRWIAMHAGCYIATSKVTLLEGMCQ 2261
QY 3087 LPRDEETNOFLWMLIEMAKOACKERKHVSLSKTCOPRSNEDNEFASELLRPGQ--N 3143
DB 2262 LP-DONTHQFLWMPREWSHPCAKRQKLENEVKREC--ASACIIEYGTIDPVCBEACT 2318
QY 3144 DIRKYSILNLKKNMENINITYKOLKODSSGNINDKRE-ENVOYSIKRSOCALEN 3202
DB 2319 QYRDYITRKIQ-----EYRLANTQYNTNFENKKEVTKAPYENDK---CNDKCN 2365
QY 3203 DINEIYTGKNNENNFEKVLKLYPGILFVDETHKNHVLGNIKEEQYTPRALYF 3262
DB 2366 CLSKYT-----DIEKKMNMDSF---DDNOLKNCICRQIKPRP---PKVKRE 2410
QY 3263 TPHVDSFYQAPLFSTHRAQYDP-----KNDILKSSISVIVSALGIALHFMKKFKS 3316
DB 2411 EERTPEEQTPPLPPKPDLPPEAPENRDLIEKTIPEFGIALALGSAFLFLKKTKS 2470
QY 3317 SV-DLIRITINIQEGMFTLESKNYIYRSGPYGKTYIMEGT---SDEDEKYMMD 3372
DB 2471 SVGNLFQIILHPRSDIDPTKLSPNRYIPTYSKRYIYLEGSGTSDTHY--- 2527
QY 3373 LSSSDIT-SESSEYELDINDIYVPSPKYKTLIEVLEPSPKRDIPSD--DTPS----- 3423
DB 2528 ---SDTSSSESEYELDINDIYVPSPKYKTLIEVLEPSPKRDIPSD--DTPS----- 2584
QY 3424 NDPTPTNFIDENNELHDFVSYTL---PTEPNNTKASADIPNTEPNITLISDNPEE 3479
DB 2585 ONDGIPSSKITDENWNTLDEFTISMLQNEPNTPE--NNMIGYVNDNTHPTT--SRANVEE 2641
QY 3480 KPRITISIHROLYTGEKISYINIMSTNTNDIPMANRDSYRGIDILINSL 3530
DB 2642 KPRIMSIHROLYTGEKISYINIMSTNTNDIPMANRDSYRGIDILINSL 2689

```

RESULT 5
AAW22482
ID AAW22482 standard; Protein; 2710 AA.

```

XX AC AAW22482;
XX DT 07-OCT-1997 (first entry)
XX DE Plasmodium Proj3.
XX DB DBL gene family; SAMP, sialic acid binding protein; vaccine; therapy;
XX KW Duffy antigen binding protein; Duffy antigen binding protein; erythrocyte;
XX KW DABP; mezoizole; malaria; var-1; var-2; var-3; var-7; immune response;
XX KW Plasmodium.
XX OS Plasmodium falciparum.
XX PN WO9640766-A2.
XX PD 19-DEC-1996.
XX PF 07-JUN-1996; 96WO-0509508.
XX PR 07-JUN-1995; 95US-0487826.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PI Chitlins C, Miller LH, Peterson DS, Sim KL, Su X;
XX PI Wellens TE;
XX DR WPI: 1997-052231/05.
XX DR N-PSDB: AAT72897.
XX PS New malaria vaccines - contains cysteine-rich DBL family protein
XX PT binding domains homologous domains of the Duffy and sialic acid
XX PT binding proteins
XX PT disclosure; Page 50-56; 96pp; English.
XX CC This sequence represents Proj3 of Plasmodium. Proj3 belongs to
XX CC the Duffy binding like (DBL) family of genes which have homology to the
XX CC Duffy antigen binding protein (DABP) and sialic acid binding protein
XX CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
XX CC var family of genes modulate cytoadherence and antigenic variation of
XX CC Plasmodium infected erythrocytes. SAMP and the Duffy antigen binding
XX CC protein (DABP) are soluble proteins that appear in the culture
XX CC supernatant after infected erythrocytes release merozoites. DABP and
XX CC SAMP mediate the binding of merozoites and schizonts to the erythrocyte
XX CC surface. These proteins are necessary for erythrocyte invasion by the
XX CC parasite. This sequence can be used in the compositions of the
XX CC invention. The compositions are for the treatment and prevention of
XX CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
XX CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
XX CC family of genes having homology with conserved regions of DABP and SAMP.
XX CC The compositions are used for the treatment and prevention of malaria.
XX CC They are also used in the preparation of vaccines for inducing a
XX CC protective immune response in a mammal to Plasmodium merozoites
XX CC (especially Plasmodium falciparum or Plasmodium vivax).
XX CC
XX SQ Sequence 2710 AA;
XX
XX Query Match 15.3%; Score 2972.5; DB 18; Length 2710;
XX Best Local Similarity 26.5%; Pred. No. 2.6e-159;
XX Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;
QY 9 IIRKGNASSLEGGAKSPITIKESHKSAANVLEKAKNIRHRSK-YAKHVDLSKGLDTKA 67
DB 6 LAKMPKRAAGDDIED-----ESAKHMDRIGKGDYDVYKKEAKRGKGLGRISA 58
QY 68 EFRGSPSTPVNKHNYYPYPCNLDEKHTNLRDYDVNLRHPCHGREGNRFDEDESEC-G 126
DB 59 KEKNESDPQPED-----PCDLDEKHTNLRDYDVNLRHPCHGREGNRFDEDESEC-G 110
QY 127 NKIRYKRRKNDIACAPRRRRMCDKNLEALNDINTONIHDLGNVLTARYEESIYNN 186
DB 111 NKIKSQGQDNKACAPRRRLHVCQNLQETPIKITHTNHLVDVCAAKFEGQSTIOD 170

```


QY 187 HP-HKGT-----SDACTALARSFADIDIVRGIDMFKPNVHD-----KVTGLREVEFKI 235
171 YPKQOAYGDSPOICTMILARSFADIDIVRGIDLYLGNFOEIKOQOOLENNILKTIKGI 230
QY 236 HD---GMEDEVKANDYNDGSGNYKILREAMNNVNRKVWVAITCDASYKGYFQMSNT 292
231 YKLNKEAARYGND--PE---FFKILREDMWTANRETVMVAITCNA--WGNTYF--HATCNR 282
QY 293 PLSNPACGKHOGKVPYTNLDVPOYLKMPDEMGCEFCRKNIKLKRYKDCSR---NDKE 348
283 GERKYGCRONDDOVPYFYFYVPOYLKMFEBMAEDFCRKNKIKKDVKRCNRGRDKEDKO 342
QY 349 RLYCSHNGDCTTITMKKGLIHLDNKCTDSTCKYKVEWLGNOQAFKKOKERYEIQ 408
343 R-YCSRNGDCEKTKRAIGLRYKQOCISCLYACNPYVDWVNNOKEOFDKOKKRYDEIK 401
QY 409 SY-----LSNDNKFVNNINSE--YKQFYEKLETOYATNDTELNLNEGKVC----- 454
402 KYENGASGSRORADAGGTTTNDYGEKFEYDELNSETRYDKFEKLSNEEICTYKVK 461
455 --KGLGPERKDIPTNSA-----DDKGLFYRSEYCOVPCDGVKCDGJKYTH--KSDND 504
462 DEEGTIDFKNNVNSDSTSASGTNVSOGTFYRKYCQPCPYCGVK---KVNNGSSNE 517
QY 505 RERYNN-----EDYKPPWGVKPTNITVLYSGNEQDITOKLEFCN-----SS 547
518 WEEKNNGCKCKSGKLYEKPPEKEGTTITILKSGKHDIIEELKFCDEKNGDTINSGG 577
QY 548 TNYKDKNN-----OKMECYKNDENINRCKLEONTEINDNPR-----II----- 586
578 TGGSGGNSGRORLEYEEMKC-YKGEDVYKVGHDEDEDEYENVNAAGLICILKNOCKKE 636
QY 587 -----SFNFEFLVWLYLNDITKMNKDKTCI--NNTTTC--IDECNRNC 629
637 EGGNTSEKPEDELOKTNPFYVVAHMLDSIMKKKLRCLONGRIRKONGKNCND 696
QY 630 LCPDRWYKQKEEWNSTKLLFTKKNKIQ-----OSYNNINNFEGYF--- 672
697 ECFKRWITOKKDEWGIYVQHF-KYONIKGRGSDNTAELIYFDHVDYLOVNLDEEFLKD 755
QY 673 -----FKYMDKLDKDAKKKELMENIKRKNKNEFSNLENNRD-----YLEVAITLL 718
756 SEDASEKSENSIDAEEA-----EELKHLREITSEDNNOEASVGGVTEOKKIMDKL 809
QY 719 DNLKETATIC-----KDNNTNEACETSHNATTNFCVPRGSGTO--PTKNIKETIQ 766
810 NYEKDEADLCLEIHEDEEKEKGDNECEIEEENFRYNCSSGSKRYRVLANKVAYQ 869
767 YFKRSATVEEARNKGLHLKGAHEGITYKRGRRKDFKDNLCRIMIKHSNNLGFSGPCD 826
870 MHHKAKTOLASRAGRSALRDISLAQFKNGRNGSTLKGOLCKIMENYSNRSNGSGPCY 929
QY 827 KGGTGDGDIOTREFVYGTMEVDPBEHMRKHEDVIMPRRRHICTSNLEHLOTDHPLNGNI 886
930 GK-DGDDGVRMARIGTMSNIEGKKOTSYKNVFLPRRREHCTSNLENDV-----GSV 982
QY 887 V-DOLVNSFLGDVLLSAKYEAANKIIMYKERNKLKGPKEVTDV---KHOTTICRAIYS 942
983 TKNDKASHSLLDGVOYLAQDAEIIIRKYDQNNI---QUTDPTQCKDEAMRAVAYS 1038
QY 943 FADIGDIIIRGRDLMERKGDVAKLOGHLEYVGNHKSLLG-KGMDKXNDV---APKTLK 998
1039 FADIGDIIIRGRDMDDEKSSNDMETRLITYEKNIKERKHDIKNDPKYTGSESKPAPAKL 1098
QY 999 RENMEANRAKVEAMKCDIKYLLKDKSGHSTOSSYSGYSDHPTLDDIYIOPKLRMTEMA 1098
1099 RADWMEANRHOVWRAKMC-----ATKIIIC---PGMVDYDIOPKLRMTEMA 1143
QY 1059 EWTCKYOKKEVDKLEKCKECKDKNDGCGTKESGTCTKCTEACNEYNIDILGKMDWN 1118
1144 EWTCKASOEYDKLKICADCMKSGDK-CT--QGDVDCGKKKACDKRYKEETIKEMNDQMR 1201

QY 1119 IISDKYELHEQOAMVSNSGIEASTAKNHID---RNVIETSELYQONGK---SNK 1171
1202 KISDKYMLTYOAKTSTNPG-----RTVLDGDDDDPYOQWDFLPIHAKSIAARVLKRA 1257
QY 1172 SGTDSBAVIGTNTYENVCAYILHDITGNFDCCOQNFCEDEK-----SOGKNEKTAFR 1225
1258 AGSPTETAAAAPITPYSTAAGYTHQEIYGCGOQOTOFCEKHKHATSTSTTKREKEYTFK 1317
QY 1226 DKPDHDGACGCKSGSRPTVQOITKTKKAEBKDEKCTVNDILKENDKKOVEDCHPKN 1285
1318 OPRPEYVTAOCDINRSQTE-----EPKKEEVEBACATVKEILLEGKNGRTTVGCNCKES 1373
QY 1286 SNGTIPMOC--GNINLVDDPRYVCPRRORCKVHFELANDNEIKLOQOVNLEKAFKSA 1344
1374 ---YPMDCCKNNIDISHD--CACHPRRORCKLCTLYIAHESOTENIKTIDNLKDAFIKTA 1429
QY 1345 ETEFSWYKYSK--DGEENELDKELBEKIPAPILRSFMYFGVDRDLDFGDISKNGEG 1403
1430 ETEFSWYKYSKNDSEKILDR---GLIBQFIRSMKYTFGDRDCLNTDISKONDV 1485
QY 1404 SKLEQOISLEKNGDQKSPNGKTRQEWMTESHEIMFAMICALVYKIGAKKD---FTENY 1460
1486 AKAKDKIGKFFSKGSKSPSGLSQEWMTKNGPEIMKGMICALVYTPDNDKRIKINDY 1545
QY 1461 GYNNVKSFK--STTLEPRAKRPFLRWLTEMWYDYCYTRQYKLDVOEKCS--NDOLKCD 1518
1546 SYDKVNOSONGNPSLEEFPAKPOPLRMWIEWGEFCAEROKENIIRDACNEINSTOQN 1605
QY 1519 T---ECKKCEDYKVKMK--KKEMIPQDKYKXDRDKKRPDRQIIGVWYDITGNATDY 1574
1606 DAKHRCQACAYOEYENKKEEFGOTNNEFVLKANOPODEPKGYEYKD----- 1656
QY 1575 LNRKFTASCGDKPSASVAVOIRIOLLEKQAYYADAKHCGCTKFEINDKYTNISSKDK 1634
1657 -----GVQF-----IGNEVYLQK---CDNNK--CSC----- 1678
QY 1635 GLVKEANTGAIKWKONKPNNTNNILKELTEDVLFPSRLRICFHALDGNVYDPEYKDNGL 1694
1679 -----MDGNVLSVSPERK--- 1691
QY 1695 RKRLMEVAATEGYLGOYKKEKKEKIKITSDAKYSYEVPPCSAMKYSFDLDDIIGI 1754
1692 -----PFGKX-----AKTYP----- 1701
QY 1755 DNLDEKQTEENLKKIFNKNGTSVSGKSDSTTGNPSTARKFPWENNEKBEVWNAITGY 1814
1702 ----- 1701
QY 1815 KRGDGNSGNSARSDEDLKCKGSPSDDDYPMGKNDEGTAYOFLRMFAEWGDEPCKHK 1874
1702 ----- 1701
QY 1875 EKELEKLVGACNDYTCGDNEDKRCCKTDACQYKKFISEMKPOYEKQIKRYGENKDIYS 1934
1702 -----EKC----- 1704
QY 1935 EHPVAKDAEDAREYLDKOLKKTICENKSGDCEYKCMKDVSTORLTDGNSQMPASLDEPK 1994
1705 -----DC----- 1706
QY 1995 EVEGCKMCOYPRGPRRRETPSPRVSLISKATASKKEAKTAPPTKOPKKEYNLTTMRA 2054
1707 -YQOK---HVPSTIP-----PPPVOPQ----- 1725
QY 2055 QTRTRAAQOTRKRTSTATTESDVGTMYKRAILSNKPSDRSGICGCPKRYGOV--PKWG 2112
1726 -----PPAPYTYDVYCSIVKTLF---DINNFSACGLK-YGKTAPSSWK 1766
QY 2113 CI-----VGKRENGEIGCMPRRKILINNIOYLVNLET-----ENKRD 2151
1767 CIPSDTKSGAGATYKSGSDGSGICIPRRRLRYVGLQ--EWATALPQGGAPSHSRA 1824
QY 2152 NDIKEAFIKCAAITOTPLMKIITE-NPAEENF-----LONGT 2188

```

Db 1825 DLIARNFISALETFFLDWDRKKEEKKFQDGSQALSLQTLSTYSDEDEDPDKLLQNK 1884
Qy 2189 IDDEFKRIYVYGYDYKDMFEGDISNDKRIITVNTIENNNKKKQDKKDEELR 2248
Db 1885 IPPDFLRMEYTLGROYDLVHGNTSDSGNTNGSNNNIVLEASGNKEDMKIQKIO 1944
Qy 2249 I-----FWEKKKFIWGMITYGLY-----HLDNENKERT 2280
Db 1945 ILPKNGTLPVPSAQTDPDKMNEHAEISIMGMICALYTEKNPDTSGRGENKIEKD 2004
Qy 2281 DNY-----QYN-DMTKL-----TPSEEVYKRP 2302
Db 2005 EYIEKFEFSTADKHGTAFTGTGYKYIYKVKLEDTSGATPSASDTPILSDVLRP 2064
Qy 2303 QFLRMFTEAEFCNKRKDLKLEAGCKEY-----CN---GSN 2339
Db 2065 PYFRYLEEFGNQFCKRKRKHLQIKHECKVEENGSGRGGITROYSGDEACNEMLPN 2124
Qy 2340 DKRTQE-----CAEACVYQNFITKKKTEYERQREKFKDKD----- 2376
Db 2125 DDTVPDLKPSCAKPCSSYRKWIESKFEFEQKAYEQQKCVNGSKHNGFCETLT 2184
Qy 2377 -GKYYKY-----PSTERDIEKATCAHEYLNMKLELCSNKCSCQMKPSQLPTTQOS 2430
Db 2185 TSSKAKDFLKTIGPCKPNVBEKTIFFDDKTFK-----HFKDCDLKTSVCKR----- 2234
Qy 2431 OSSDANDMPESLDYVPEEFNKCEPBLSKSGSMITKTKTEKIPMNCVEKAAVYLSKEA 2490
Db 2235 -----DECD-----NSKGDCRKK-----NSID-----ATDI 2256
Qy 2491 ENNMDTLKEKFIPIESTREKSKNSMTNNPCDPRKPYAPKTYGRNRCEN----- 2543
Db 2257 ENGCVSTV-----LEMRVASDSKSGFNCD-----GLENACGAGIFECI 2295
Qy 2544 REENFEKVDYEMKCYKNSKFTYQEKRCVCPPRRHHCLNLDIEKIERLKSQNYLLKMYR 2603
Db 2296 RMD-----EMKC-RNVCGY-----VCKP----- 2313
Qy 2604 RTARNEGIDITIKNFENSENCAMNPICDTMKYSFADLGIIVGTMLRIGVLPVEIKLY 2663
Db 2314 -----ENVNGE-----AKGKHIIQIALVK-----RMV 2336
Qy 2664 RVEEYIYGRKRNKNGKKNYND-----VOTFRSAMWDMANKDITKAMTCAPEDAKLFRG 2719
Db 2337 EYFEEDYKIKIKIRIKNGISPCINKCYKAWDQKRKE-WKEITTEFKD----- 2387
Qy 2720 RMDGEFRTLLQDKCGHKDDPEVDYIPQRFKMTWSEYCYKALMELEKSKSCDHCK 2779
Db 2388 -----QYKNDNSDDDNVRSFLETLIPQITDAN---AKNKVITLSKSGNSCG-CS 2432
Qy 2780 TSDRKNYIDENKCKOCTRCOEYKNEFYALKKSLFDIOSNKYKELYEODIYKISTYDHV 2839
Db 2433 AS-----ANQONKN-----GEYKDAI-----DCMLKTKLD-----KIECECK 2465
Qy 2840 QNFVOKLTFKSECVSESEFSEYLHETSKLANKFNENDGSSIRIYAFEEPKSYKE--- 2896
Db 2466 HH-----QTSDECSDTQPOPLED---ETLDDITETRAKNNMMPKIGENVIKTAQODE 2518
Qy 2897 -ACSTLPDSKPNLDCPTDQKD 2918
Db 2519 GGC---VPAENSEEPAADSGKE 2538

```

```

XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Stialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
KW protozoacide; Proj3.
OS Plasmodium falciparum.
XX US593827-A.
PN 30-NOV-1999.
XX 07-JUN-1995; 95US-0487826.
XX 10-SEP-1993; 93US-0119677.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX Sim KL, Chlunis C, Peterson DS, Su X, Wellem TE, Miller LH;
XX WPI: 2000-194198/17.
XX N-PSDB; AA298286.
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
XX falciparum erythrocyte binding proteins useful for vaccinating against
XX malaria -
XX Disclosure; Columns 79-92; 93pp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
XX Binding Protein (SABP), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABP are the
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
XX and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX Immunization with the polypeptide provides effective protection against
XX malaria. The present sequence represents the Proj3 binding domain
XX polypeptide.
XX
XX Sequence 2710 AA:
XX
XX Query Match 15.3%; Score 2972.5; DB 21; Length 2710;
XX Best Local Similarity 26.5%; Pred. No. 2.6e-169;
XX Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;
XX
XX 9 IIKMNAASSLEGDAKSPITIKESHKSARNVLERAKNIRHSK-YAKENHVSIGDLTKA 67
XX 6 LAKMGPKNAAGDIED-----ESAKMFDRIGKDYDKYKEBAKERGKGLGRSEA 58
XX
XX 68 EFRGSPSTPVNKHNYYPNCLDKERTNLRYDDVNLRHPCGHGSEONRPFDEDESEC-G 126
XX 59 KFEKNESEDQTPED-----PCDLLHKYHTV---TTVINPCADRSVRFDEYGGCCTH 110
XX
XX 127 NKIRYKKNKNAIACAPRRRHMKDKNLEALNDINTQIHLLGNVLTAYABESLYNN 186
XX 111 NRIDSQGDGDKACAPRRRLHVCDDQLEIPIKITHTNMLLDVCAAFKFEOSITOD 170
XX
XX 187 HP-HKGT-----SDACTALARSFADIGDIVGIMKFNKVNVD-----KVEGLREVFKI 235
XX 171 YPKYQATYGDSPOICTMLASFPADIGIVGRDLYLGNPEIQROROLENNLTITGKI 230
XX
XX 236 HD---GMEDEVKNDYNPDGSGNYKLRBAWNNVNRKNVKAITDASYSKSGYFMSHST 292
XX 231 YEKINGAARFGND--PE-----FFKLREDWTANREYWKAITQNA-WGNYYF-HATCNR 282
XX
XX 293 PLFSNPKCGHKGQVPRYLDVYPOLYLRNFDWGEFCKRNKIKLKYKDSGR---NDKE 348
XX 283 GERTKGYCRNDQVPTFYDFVYOLYLRNFEEMADDFCKKKNKIKDYKRNCRGDKDEKD 342
XX
XX 349 RLYCSHNGHDCCTTIWKGILHLDNKCDCSTKCKVEVWLGNOAEFKKQKKEKYEKEIO 408
XX

```

RESULT 6
 ID AAY77904 standard; Protein: 2710 AA.
 AC AAY77904;
 DT 13-JUN-2000 (first entry)
 DE P. falciparum Proj3 binding domain polypeptide.

Db 343 R-YCSNRNGYDCETKRAIGKRGKOCISCLYACNPIYVDMINNOKEOFOKKKYEIEK 401
 QY 409 SY-----LSNDKFNANNINSE--YKQFYEKLEETOYATDTELNLNNEKRYC---- 454
 Db 402 KYENGASGSGRQKRDAGGTTTNTVDYGEKKFYDELNKESEYRVADKLEKLSNEIECTKVK 461
 QY 455 --KGLPGEEDITFTNSA-----DDKGFYSEYCOVCPDGCYCDGKIKYH--KSDND 504
 Db 462 DEEGGTTDFKNVNSDSTSGASGTNVEOQGFYRSKYCQPCPYCGVK---KVNNGSSNE 517
 QY 505 RERNVN-----EDYKPPMGVKNPTNITLYVSGNODGDIOTKLEFNCN-----SS 547
 Db 518 WEEKNNGKCKSGKLYEPKPEKEGTTITILKSGHDDIEKLNKFKCEKNGDITNSGSG 577
 QY 548 TNYKDKNN-----OKMECYRDEINRCKLEONTLEINNDNR-----I----- 586
 Db 578 TGSSGSGNSGRQLEYEBC--YKGEDVYKVGHDEDEDEYENVNAGLICILKNOKKNE 636
 QY 587 -----SFHNFELMYTLLDITKNDKLTCT--NNTTTC--IDEENRC 629
 Db 637 EGGNTSEKEDDEIOKTFNPFYVVAHMLKDSIHMKKLOLQNGNRIRICGNKKNND 696
 QY 630 LCEPRAWKOKKEEMNSIKLFTKKNKNO-----OSYENINNLFGYF--- 672
 Db 697 ECKKRNITQKKDEKTIYOHF--KTQNTKGRGSDNTAELLIPDHDIYLOYNLODEEFLKD 755
 QY 673 -----FKYMDLKDEAKKELMENIKRKNESNLENNRD-----YLENAIELL 718
 Db 756 SEDASEKSENSIDAEBE-----ELKHLREIIESEDNNOEASVGGVTEOKNIMDKL 809
 QY 719 DHIKERTIC-----KONNTNEACETSHNATINCVCYPRGTO--PTKNIKELAQ 766
 Db 810 NYKDEADLLEIHEDEEKEKGDGNECTEEGENFYKNSGSESGKRYPVLANKYAQ 869
 QY 767 YFRRSAYEERANGLHKLKKAHEGTYKRGRRKDFKDNLCRIMIKNSNLFSGNSPC 826
 Db 870 MHRKATQULASRAGRSLRDLISLAQFKNGRNSITLKGQCKINENSNDSRSGSPCT 929
 QY 827 GKGTGDIOTRYVYGVTEWEYDEPHMKRKHEDVJMPRRRHICTSNLEHLOTDDHPLNGT 886
 Db 930 GK--DGHGGMRYMIGTEWMSIEBKQKQTSYKNVFLPPRREHCTSNLEHNLV-----GSV 982
 QY 887 V--DVLVNSFLGVLVLSAKYEANKIITRMKKNLNGKPRKVTDP---KHOTTICRATYS 942
 Db 983 TKNDKASHSLGVOQLAAKDAEILIKRYKDNNT---OLTPLOOKDDEANCRATYS 1038
 QY 943 FADIGIIRGDLERNGDVAVKLOHLETFVGNIHKSLG--KGNDKYND---APRYKL 998
 Db 1039 FADLGDIIRGRDMDEKSDTMDETRLITYFKNIKKEHDSIKDNPKYTGPESKRPAYKL 1098
 QY 999 RENMMTEANRAKVAEMKCDIKYILKNSGHOSSTOSSYCGYSDHPLDDYIPQKLKMTEMA 1058
 Db 1099 RAOMWMEANRHOVRAMKC-----ATKGIIC---PCMEVDDYIPQRLKMTEMA 1143
 QY 1059 EWTCKYOKKEEYDLKECKECKDKNDNGOGCTKSGTCTEACENXYNDIIGLMEKQWN 1118
 Db 1144 EWTCKKQSOEYDLKTKICACMSKSGDK--CT--GQVDVCGKCAACDKYKEIEEIMWDMR 1201
 QY 1119 IISDKYELHQAOMYSVNSGIEASSTAKNHID---RNVLEFLSELQOONGK---SNK 1171
 Db 1202 KISDKYVLTLOAKTSTNG-----RTVLGDDDPDYQOMVDFLPIHKAISIAARVLYKRA 1257
 QY 1172 SGFSDESAVIGTYTTEVNGAVYLDHGNFPDQSONEFCEK-----SGKDNKAYFR 1225
 Db 1258 AGSPETIAAAPTTPSTAAGYHOEIGYGOEOTOFCERKHGANTSTYTKKEKEETFK 1317
 QY 1226 DKFODHDGAGCKSGSFTVJOITKKAKEKDEKTEVDILKENDGKGOVEDCHCKN 1285
 Db 1318 QPPEVYATACDINRSTGTE---EPKKKEENVESACIYKELIEGKNGKRTTVEGCNKES 1373
 QY 1286 SNGPYDMQC--GNINLVDPKVCMPRRQKLCVHEFLANDNEIKLOSOVNLKEAFIKSAA 1344
 Db 1374 ---YPMDCCKNNIDISHD--GACMPPRROKLCVYIAHESQENIKTDNDLKDAVITKAAA 1429

QY 1345 ETEFSWYTKRSK--DGEENELDKELKEGKIPPAFLRSNFTYFGDYRDELFGTDISKRGEG 1403
 Db 1430 ETEFLSMQYKRSKNDSEKALIDR-----GLIPSOFLRSNFTYFGDYRDELFGTDISKRGONDY 1485
 QY 1404 SKLEQDLSLFKNDKOSPNGKTRQEWTEHSEHETWAMCALVKAIAKADD---FTENY 1460
 Db 1486 AKAKDITGKFFSKDGSSPGLSRQEWMTKNGPIMGMICALKRYTDTDNKRRIKNDY 1545
 QY 1461 GYNNVKSDEK--STLEEFKAPPOFLRWLTWEDYCYTRQKLYADVOEKKS--NDOLKCD 1518
 Db 1546 SYDVNOSQNGNSPBLEFAKPOFLRWLTWEDYCYTRQKLYADVOEKKS--NDOLKCD 1605
 QY 1519 T---ECNKKCEDYKYNK--KKEMIPQDKYKXKDEKRRKREFROHIGVATDYGTFNATDY 1574
 Db 1606 DAKHRCNQCRAVQOEYENKKEKFSGOTNNFVLKANVQPODEYKGYEYK----- 1656
 QY 1575 LNKFTASCGDPRSASVQNRITOLEKQAYVADKRCGCTKFTENDKTNISSKDKC 1654
 Db 1657 -----GVQP-----IGNEYLQK--CDNNK--CSC----- 1678
 QY 1635 GLVKEANTGAIKQONKGNVNNLKELETDVLFPSRLRICFHALDGNYTDPVEYKDEGL 1694
 Db 1679 -----MDGNVLSVSEK----- 1691
 QY 1695 RKRLMEVATGYNLGOYKREKEREKIKTSDAHKYSYEVPPCSAMKYSFYDLRDIILGI 1754
 Db 1692 -----PFGRY-----AHKYP----- 1701
 QY 1755 DNLEDEKQTEBNLKKITFNKNGTSVKGSDSTTGNPSTARKFTWENKECVWNAAMICGY 1814
 Db 1702 ----- 1701
 QY 1815 KRGRDNGNSARSADBDLKKCGSVPSDDDYPMKNDDEGHAQOFLRWFAWGEDPCKHK 1874
 Db 1702 ----- 1701
 QY 1875 EKELEKLVACNDYTCGDNEDKRRKCTDCTOYKFIYSEWPKOYKOKIKYGENKDIYS 1934
 Db 1702 -----EKC----- 1704
 QY 1935 EHPVAKADEAREYLDQOLKIKCNKSGDCEYKCMKDYSTQRLDNGSQQNMPASLDEPK 1994
 Db 1705 -----DC----- 1706
 QY 1995 EVEKGCNOVPRGRPRVRRTPSPRVSLSKATASKKEAKTAPPTKPKVENLTTEMA 2054
 Db 1707 -YQK--HVPSTP-----PPPPVQ----- 1725
 QY 2055 QTRFRAAOOTRKRTSTATTESDVGTWYKALISNKPDSRGIEGNCNPKTYGY--PRWG 2112
 Db 1726 -----PEAPVTVYVCSIVKTLFR---DTNNSPDAGLK--YGKTAPESSWK 1766
 QY 2113 CI-----YKSKENENGICMPRRKRLKINNIOYANTET-----ENKRD 2151
 Db 1767 CIPSDTSGAGATTGKSGSDSGSCTIPRRRRRLVYGLQ--EWATYALPQGBAAPSRA 1824
 QY 2152 NDIKAFITKCALETQFLMKLYIE--NPAENE-----LQNGT 2188
 Db 1825 DDLNNAFIOGASALIEFLMLBRYKREKRPQDGGSOQALSQLSTYSDEEDPDLQNGK 1884
 QY 2189 IPDEFKIMTYTYDYDMDFGDISNDKTIITYTNSVTITLNNKKKODKKDEELRK 2248
 Db 1885 IPPDFLMEYTYLDYDILVHGANTSDSGTNSNNNIVLEASGNKEDMOKIOEKIEQ 1944
 QY 2249 I-----FWEKNNKFTIEGAIYGLTY-----HLDEENKEXIR 2280
 Db 1945 ILPANGSTPLVPKSSAOTPPDKMWEHABSLYKKGICALITYTEKAPDTSARDEKIEKOD 2004
 QY 2281 DNY-----QYN--DMTKL-----PSTSEEFVYKRP 2302
 Db 2005 EYERKFFGSTADKRGTASTPTGYTKQYDEKXVLEDTSGAKTPSASSDPPLSDVFLRP 2054

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
XX
DR WPI, 1995-123427/16.
XX
XX N-PSDB; AA083529.
XX
PI New erythrocyte binding domain polypeptide(s) - isolated from
PI Plasmodium binding proteins, used in diagnosis, treatment and
PI prevention of malaria
XX
XX
PS Disclousure; Page 61-65; 81pp; English.
XX
CC Erythrocyte binding ligand (EBL) family genes were cloned from
CC P. falciparum chromosome 7 subsegment libraries constructed during
CC genetic studies of the chloroquine resistance locus. The 4 genes,
CC EBL-e1 (AA083526), E31a (AA083527), EBL-e2 (AA083528) and Proj3
CC (AA083529), encode the proteins given in AAR0233-36, respectively. The
CC binding domains of such proteins can be expressed e.g. in E. coli.
CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
CC cells, and provide protection against P. falciparum.
XX
XX
Sequence 2703 AA:

Query Match 15.3%; Score 2972; DB 16; Length 2703;
Best Local Similarity 26.6%; Pred. No. 2,7e+169;
Matches 861; Conservative 438; Mismatches 864; Indels 1076; Gaps 119.

33 KSAANVLERAKNIRHSK YAKEHDSLKGDLTKAEFGSGSPVKNKINYYPCND 91
16 ESAHMHMRIGKGVYDKVKEAKEERKGLGRSEAKFFKNEBDPQTPED-----PCDD 70
92 HKEHTLRVYDVALRHPCHGREGQRFDEDESEEC-GNKTRNTKRNKDAIACAPRRHMC 150
71 HKYHTNY---TIVNIPCADRSDYRSDESGGCTNKRKDSQGDNKACAPYRLHVC 127
151 DKMLEALNDINTQNIHDLGNVLVTKYGEISTVNNHP-HKGT-----SDACTALARSPA 204
128 DQMLEQIEPKITNTNHLNLYVCMAKKFGOSITODYPKQATYGDSPSQTMLARSPA 187
205 DIDIDVIRGIDMEKPNVHD-----KYVTGIREYFKTHD-----GMEDEVKNDYDPSGGNTY 256
188 DIDIDVIRGIDMEKPNVHD-----KYVTGIREYFKTHD-----GMEDEVKNDYDPSGGNTY 256
257 KLRKAAWNNVRNKKWEALITCDASVYSGYFMOSNTPLFSNKGCHGKQKYPNTLDYPO 316
242 KLRKEDWNTANRETYWKAITCNA-WGNTYF-HATCNRGERTKTCYCRNDQYPTDYDYO 299
317 YLRWPEFMEEECRKRNIKLKVKVNSCR-----NDKEBRLYCSINGHDCITTTWKGIHLMD 372
300 YLRWPEFMEAEDECRKNNKIKYKRNKRGKDEKDR-YCSBNGYDECKTRAIKGLRYG 358
373 NKCTDCTSKCYAEVWLGNOQAEFKKREKYEKEIOSY-----LSNDNKFVNINNS 423
359 KOCISLACNPNYVWINNKKOEQFKKKYKDEIKKYENGASGSGSRQKRAAGTITTY 418
424 E-YKFOEYKLEKQYATNDPFLNLNLSGKC-----KGLPGCKDITTTNSA----- 470
419 DQYEKKFTDELKSKYRYVDLFLEKLSNEEITCYKDEEGGTTIDFKVNNSDSTGASGTN 478
471 -DDKGIPTFRSEYQVPCDGVKCDGIKYTH-KSDNDRERVYN-----EDKPPWGVKPT 522
479 VESQGTFRSKYKCGCPYCGVK-----KYNNGSGSNEMEKKNKGCKGKLYEPDPDEGT 534
523 NTVIYASNEGGDITQKLEFCN-----SSTNYKDKNN-----QKWECYKRD 564
535 TITILKSKGHDDIEKLNKFCDEKNGDTLNSGSGTGSOGSGSGHQELIYEKWK-YKG 593
565 ENINNKCKLEQNTINNDPK-----II-----SFFNPELW 595
594 EDVAVGVHDEDEDEYENWAKGGLCLIKNKKKKEGGNTSEKPEDEIQTINPFPEYW 653
596 VTIIIRDITKNNDKLKTCTI-NTTTHC-IDECNNCLCFDRWVYKQKEEWNSIJKLFTTK 653

Db 654 VAHMLKDSJHKKKLOKLOLGNRIKCGNNKCNDCCECFKRWITQKDEWKTQHF-KT 712
 QY 654 KNIQ-----QSYNNINMLFEGYF-----FKYMDKLDEAKKELM 690
 Db 713 QNKGKRGSDNTAELLIPFDHDIYQVNLQEBFLKDESDASEEKSSEMSLDAEFA----- 766
 QY 691 ENIKRKNEFSNLENNND-----YLENAILLELDLHLEKATATC-----KDN 731
 Db 767 EELKHLHEIIESEDDNNOEAVSGGVTEQKINIMDLNLYEKDEADLCLEIHEDEEEKK 826
 QY 732 NTNEACETSHNATNTPCVKRGQTQ-PTKNIKELAQYFKSAVEANRNGIKHKKGAHE 790
 Db 827 DGNECIEGEGEFNRYNPGSGSGNKRYYLANKAYQMHKAKTQLABRAGSALRGDISL 886
 QY 791 GIYKRGRRRDFKNDLCRIMIKHSNMLGFSNCPDGKGTGDIOTFRVYVTEWEVPEH 850
 Db 887 AOFKNGKNGSLKQOICKINEMYSNDRSGNSGFCYCK-DGDHGVYMRIGTSEMSNIEGK 945
 QY 851 MKKDHEVIMPERRRHICTSNLEHLQTDHPLNGNIV-DDLVNSFLGADVLLSAKYBANK 909
 Db 946 KQSYKKNVFLPBRREHCTSNLENDLV-----GSVTKKNDKASHSLGADVQLAAKTDAE 999
 QY 910 IIRYKREKNLKGKKEVTD-----KHOTICRAIFRSADIGDIRGSDIMERNDMKTLO 966
 Db 1000 IIRYKQONNI-----QLTDPLOQDOEAMCRVAFYSFADIGDIRGSDIMDEKSTJME 1055
 QY 967 GHLETVGNLHKSILKG-KGNDKYND-----APKYLKLRENMEANRAVYWEAMKCDIYK 1022
 Db 1056 TRLTIVKKNKEKHNDGJKDKPRTYGDSSKKPAYKLLRADWEANRHQVYKAMKC----- 1109
 QY 1023 DKSHOSTOSSYCGYSOHTPLDITIPOKLNMTEMAEMYCCKVQKEKEDKLECKECKD 1082
 Db 1110 -----ATKILIC-----GPMADYDIPQRLRMTEMAEMYCCKVQKEKEDKLECKECKD 1160
 QY 1083 DNGOGCTKESGTCTKTEACNEYNNDIIGLMEKQWNIISDXYKELHQAQMSVNSGIEA 1142
 Db 1161 GOGK-CT-QGDVDCGKKAACDKKEIEKKNEMWRKISDXYKELHQAQMSVNSGIEA 1215
 QY 1143 SSTAKNHD-----RNVEIFELBYOONGK-----SNKSGTDESANVIGTNTYVNGAYLH 1195
 Db 1216 -RTVLGDDDDPYQOMWDFLPIHKAISTAARVLYKRAAGSPLEIAAARPIFYSTAAGI 1274
 QY 1196 DTGNEFDCCOSONERCDK-----SDGKNEKIVAFROKPODHGACGCKSGSKRTYQIK 1249
 Db 1275 QELGYGCGOBOTOPCEKKGATSTYKTKENKXYTFKQPPPEYATACQINSQTE-----E 1330
 QY 1250 TKKAKEEDKECTVNDIILEKNDGKQVODCHPKKNSNGYVPMQC-GNINLYEVPYCM 1308
 Db 1331 PKKEEVESCKTVEKILBKKNGRTVGCNPKES--YPMDCKNNDIDISHD-GACMP 1386
 QY 1309 PRROKLCVHFLANDNEIKKLOSOVNLKAEATKSAAEFTFSWYYSKSK-DEGENDELKEL 1367
 Db 1387 PRROKLCITYAHESQENIKITDNLDAFKTAAEFTLSMUYKSKNSEAKITDR-- 1444
 QY 1368 KEGITPAFLRSMYTTGDRDPLFGDISKSGHSGSKLEQIDSLKRNQDQKSPNGKTR 1427
 Db 1445 --GLIPQPLRSMYTTGDRDPLFGDISKSGHSGSKLEQIDSLKRNQDQKSPNGKTR 1502
 QY 1428 QEWMTESHEIWEAMLCALVYIGAKKDD-----FTENYGNVNYKFSDK-STILEEAKRPOF 1483
 Db 1503 QEWMTKTPRILKGMCLALRYTDTDNKRKIKNDYSDKYNQSONGNSPSTLEEPAKPOF 1562
 QY 1484 LRMWLEMYDYCYTRQRYLADVOCKS-NDOLKCDT-----ECNKRCDYVYKYM-KRKEM 1538
 Db 1563 LRMWLEMYDYCYTRQRYLADVOCKS-NDOLKCDT-----ECNKRCDYVYKYM-KRKEM 1622
 QY 1539 IPQDKYKDERDKKRPORHIGVAVTYDTGNATDYLNRKFTYACGDKPDSASVYQRIQ 1586
 Db 1623 SGQTNNEVLAAVQPOPEKGYEKD-----GVQP-----IQONEY 1659
 QY 1599 LLEKQAVYDADKHCCTKFTENDOKTYNISCKDKGLVKEANTGALIKWONKGPNNNTNL 1658

Db 1660 LLOK-----CDNNK-CSC----- 1671
 QY 1659 KETLEDVLFPSRRRLICFHALDGNVTPDEYVDENGELKRLMEVATEGYNLQYKKEKE 1718
 Db 1672 -----MDGNVLSVPREK-----PRGKT----- 1689
 QY 1719 KETIKTDANKYSTEYVPCSAKMSYFYDLRDIIGIDNLEDEKOTEBNLKIPNNKTS 1778
 Db 1690 -----AHKYP----- 1694
 QY 1779 VKGSDSTIGNPGSTARKFENNEKBEVWAMIGYKRGDRDGSNGSARSDELLKCGS 1838
 Db 1695 ----- 1694
 QY 1839 VPSDDDYPMGNRDEGTAYOLFMRFAEMGEDFCFKKEKELEKIVGACNDYTCGNEDRK 1898
 Db 1695 -----E.1695
 QY 1899 KCTDAGTQYKKFISEMPPQYEQIKTYGENKOKIYSEHPAKDAEDAREYLQKICE 1958
 Db 1696 KC----- 1697
 QY 1959 NKSQDCEYKCMKQVSTQRLTDGNSQNMPSASLDEPKVEYKCGNQVGRPPRVARETSP 2018
 Db 1698 --DC-----YQK--HVSIP-- 1710
 QY 2019 RVSLISKATASKKEAKTAPPTQOPKRYENLTETEMRAOTRTTAAQOTRKRTSTATTESD 2078
 Db 1711 -----PPPVQPO-----PEAPTYVD 1727
 QY 2079 VGVWKAIIILSKPDRSGIEGKCNPKTYQY--PWGCI-----VKSXENENG 2125
 Db 1728 VCSIVTKTLF--DTNNSDAGLK-YGKTAPSSMKCIPSDTKSAGATGKSSSDSGSI 1783
 QY 2126 CMPRRKKLCINNOYANET-----ENKRDNDIAFTKCAIEFOFLMYII 2175
 Db 1784 CIPRRRLRYGKQ--EMATLPOGEGALPSSHRADDLRAAFQSAIETFFILMDRYKE 1841
 QY 2176 E-NPAENE-----LONGITPDEKRIYVYVYGDYKDMFPTD 2212
 Db 1842 EKKPOGDSQALSQLNSTYSDEDEPPDKLQNGKIPPELRLMFTYLGDRDILVHG 1901
 QY 2213 ISNDKIIITYNSTTILNENKKNKODKKDEELKTI-----FWE 2232
 Db 1902 NTSQSGMTGNSNNNVIYLEASGNKEDQKIOEKLEOILPKNGGTPLYPKSSAQTPDKWN 1961
 QY 2253 KKKFIEGMYGLTY-----HLDENEKEXIRBNY----- 2283
 Db 1962 EHAESIMKMCALITYTEKNPDTARSQDENKIEKDEYERFPGSTADKGTASTPTGTY 2021
 QY 2284 --QYN-DMTKL-----TPSLEEFVVRPQFLRMFTMAEFCNKRREOLIKL 2326
 Db 2022 KTYQDYERKVLIEDTSGAKTPSASDTPLLSDVFLRPPFRVLEMGOFCKRRHKLQI 2081
 QY 2327 EAGCKEYF-----CN--GSNDKTOE--CAEKCVTYQNF 2358
 Db 2082 KHECKVEENGGSRGGITROYSGDGEACNEMLRKNDGTVPDLKPSCAKCSYSRWIE 2141
 QY 2359 KWKTEYEROREKFKKRD-----GKKYKDY-----PSTEDEIKAT 2394
 Db 2142 SKGKEFEKOEAYQOQDKVCYNSKNKHDNGRCETLITSSRAKDLTKLGPCKPNNVEKTY 2201
 QY 2395 CAHEYLNKLIKELGKNDCCSMQRPSSOLPRTQOOSOSDANDMPESLDVPPEEFNCEC 2454
 Db 2202 IFDDOKTFK-----HFKDCDCLKFSVCK-- 2231
 QY 2455 PELSKGSMITHKTIETPKIPMNCVEKAAAYLSAEANMDITLKEFPIESTREKESK 2514
 Db 2232 --NSKGTDCRNK-----NSID-----AVDIENGVDSTV--LEMRVSADSK 2267
 QY 2515 NSMTNNNPNCDPKRPYADKYIGRRNPNEN--BEENEFKVDYEMKCKYKNSKPFQEK 2567
 Db 2268 SGFNGD-----GLENNACRGAGIFEGIKRD-----EMKC-RNVGTY-- 2302

QY 2568 KRCVPPRRHMCILNDEIKERLKSNNYLLKMWRTARNEGIDITKNSENGCAMNP 2627
 Db 2303 --VCKP-----ENVNGE----- 2312
 QY 2628 ICDIMKKSFAIDGIVKGTDMRLRGVLPVEIKLYVFEYIKWNRKNGKRYND-- 2665
 Db 2313 -----AKGHIIQIRALVK-----RWVEFFEDYKIKHKKISHRIKNGEIS 2353
 QY 2686 --VQFRSAMDANRKDIWKAMTCAPEDAKLFKRGHMDGERTLLDQCGHDDPPVD 2743
 Db 2354 PCIKNCVKEWQDKRKE-WKEITFRFD-----QYRNDSSDDNVR 2393
 QY 2744 DYIQRRRWMTWSEYCYKALMBELFKFKSCHCKTSRCKNDYDNCKECCQTRCOEY 2803
 Db 2394 SFLETLIPQITDAN--AKKNVILSKFGNSCG-CSAS--ANEQNNK-----GEY 2437
 QY 2804 KNEFLAKKSLFDIOSNKKELYEQPIYTKISTYDHQNFQOKLKTFFSECSSESFEYLH 2863
 Db 2438 KDAI-----DCMLKALKD-----KIGCECKKH-----QTSDECSDPPOQPLE 2477
 Db 2864 ETSKCLMYKENENDGSSNIRTYAFETPKSYKE---ACSTLPKKNPLDNCPTDQMKD 2918
 Db 2478 D--ETLDDIETBEAKKNMPPKICENVLKTAAQDEBGGC--VPAENSEPAAITDSGE 2531

RESULT 8

AA18352
 ID AA18352 standard; Protein; 2197 AA.

AC AA18352;

DT 07-NOV-2000 (first entry)

DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:210.

KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;

XX antimalarial; malaria; protozoa; infection; insecticide.

OS Plasmodium falciparum.

PN WO200025728-A2.

PD 11-MAY-2000.

PF 05-NOV-1999; 99WO-US26796.

PR 05-NOV-1998; 98US-0107131.

PA (HOFF/) HOFFMAN S.

PA (CARU/) CARUCCI D.

PA (GARD/) GARDNER M.

PA (VENT/) VENTER J C.

PI Hoffman S, Carucci D, Gardner M, Venter JC;

DR WPI; 2000-365347/31.

PT Proteins encoded by chromosome 2 of the human malaria parasite,

PT Plasmodium falciparum, useful as antimalarial vaccines and in the

PT diagnosis of P.falciparum infection -

PS Disclosure; Page 441-447; 577pp; English.

CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AA18144 to AA18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.

QY Sequence 2197 AA;
 Db 33 KSARVLEBYAKNIRHPSK-YAKEHVDLSKGLTAEFRGSGSTVYKHNHYPPCND 91
 Db 12 ESKHMFEDIGEDYEQVSEYVSELEKLSLAPILGVSGSTNE-----TCMLV 64
 QY 92 HKEHTNLRYDYNLRHPCG---REQNRDEDESECGNKI---RNYKRKN--DAIACA 142
 Db 65 QDYNNKPYGNSN-RYPCNKILKGINERFSDTLGGQCNKIKNGEYSGKDCGACA 123
 QY 143 PRRRRMCKNLEALNDINTONIHDLGNVLTARYEESIVNHP-HKGT-----SDAC 196
 Db 124 PYRLHLSHNEISID--TTSWTHKLLLEVCAAEKESIDTHYQORINEDSPQIC 181
 QY 197 TALASFADIGDVGIDMEKPN-----VHDKVEGLREVKIHDGMEDE--YANDNP 249
 Db 182 TMLASFADIGDVGKDLFYNSKEKEKDELENTLTITKHEKLDKDEGAEFRGS 241
 QY 250 DGGSYTKLREAWNVNRRKWEALITCDASYKSGYFMS---ESNTPLESNPKC----- 300
 Db 242 D--TTYLYOLREDWYANATWEALITCDV-HGSDYFRQTCGDKETATRVAKDCKRDEN 299
 QY 301 GHRGQ---KVFNDYPOYLRFWDEWGEFECRRNKLKLVKDSCKNDKERLYCSNG 356
 Db 300 GKPPSNADQVPTDYDYLRFMEAEDEPCRRKKKLEKLDQCCYKONLYCSNG 359
 QY 357 HDCTTTKKGILHDKNCTDCTCKVFEVWLNQDPAFKKREKYEKEIOSYLS----- 412
 Db 360 YDCTIYKKGKLYGHECTJMSVCRLESIDNOKLEFLKOKKETELESNGSGCGS 419
 QY 413 -----NDNKFVNNINSE-----YKOPYEKLEQVATNDTFLNLNEGYCKGGLPGE 461
 Db 420 GGVGRNRRKKGAGVETATNTDGYEKKFYELKSEYKGVDDFLKLNNEDVCKIRDEK 479
 QY 462 KDITFTNSAD---DKGIFRSEYQVCPDGVACDGIKTYHKSNDNRRERNNDYKPPW 517
 Db 480 EKIDFTPADKNSNNEGFFHSEYCKPCPDGVARKDNQMKDKDGCCTR--GLYEPAS 537
 QY 518 GVKPNTIVLYSGNQGDIQKLENFCN---SSTN-----YKRNOK-----W 558
 Db 538 GAQGPPIKILKSGEKEIETKLFKAFDQITNGDTNSVARGADGSGKSNSRELYEW 597
 QY 559 EGYRK-----DENINR-----CKLE-----QNTFINNDPKIISFH 589
 Db 598 KCYNEVQKVDKNGEEDDEEDVDYKKAAGCICILENNKHESRNNSSNEPPOFOTFH 657
 QY 590 NEFELWYTYLLRDIKWNKDKLKTICNN-TTHICIDEENRMLCFDRVAKOKEEPMNSIKK 648
 Db 658 DFFYFWIGRFLNDSMYRGVNCINNPKRKKCKNECKDCCGCFKWKIGKKKEWEIKK 717
 QY 649 LFTKKNNIQSYVINNNLEBGFYFKVMDLKDDEAKWKLMEIKRKKNEFSLNNRD 708
 Db 718 HF-----KTQAFARN-KRENSGIMFSGIMD 742
 QY 709 YLENAIELLDHLKETATICKDNNTNEACETSHNATPCVVRGQGPYKKNKEIAQVF 768

Db 743 SADVLELAL-----ELEOLF 758
QY 769 KRSAYEARNRGLHLKGAHEGIYKRGGRKDFKNDLCRIMIKHSNRNLGFSNGPCDG 828
Db 759 -----QDIKO-----763
QY 829 GTGDCIOTFRVGT EWEVDEHMRKCHEDVYIMPPRRRHICTSLNLEHLOTDDHPLNGNIYD 888
Db 764 GYGD-----VKELKGIKELDEEKKKQAEAVV-----VV-794
QY 889 DLVNSHSLGVDLLSAKYEANKIIRMYKEKNLNGPKREVDPKHOTTCRAIRYFADIGD 948
Db 795 -----VADNOKTITDILLOHE-----811
QY 949 IIRGRDLMERNGDVYLOGLHLETFVGNHSLKSGKNDKYNDAPKYLKLRNWEANRA 1008
Db 812 -----GDDANNCLKTH-----822
QY 1009 KWEAMKCDIKYLDKSGHSTOSSYCGSDHTPLDDYIPQKLRMTWMAEWYCKVQOKE 1068
Db 823 -----822
QY 1069 YDLKKECKECKDKD--NGOGCTKESGCTGCTKCTACNEYNIDILGKWEQNNIISDKY 1125
Db 823 -----KEKCEOTQKRPAGAGPGAPSETGTT-----LEDEEE 856
QY 1126 ELHEQAOMSVNSGIEASSTAKNHIDRNVIETLSELYOQNGKSNKSGTDESAVIGTNT 1185
Db 857 EEDEE-----EDAGDEV-----EEGETVDT- 876
QY 1186 TYENVGAYLHDTCNFDCCOSONEPCDEKSPDKONEKAYAPDKPODHDGACGCSGSKPTR 1245
Db 877 -----TEGDETEVEQVFKDJDREG-----896
QY 1246 VOJTKKKAEEKDE--CKTYNDILKENDGKKQVEDCHPKKNSNG--YPMOC--GN 1296
Db 897 -EEBEAKKATDTTSLDYCOTVKNALNND--NLTDACKLKYGBGGERPPMKCVSSGE 953
QY 1297 INLV-----EDPRVCMPPRRROKLVHFLANDNEIKKLOSOVNLKEAFIKSAA 1344
Db 954 KSVATGSSGATGSGDKGALCVPRRRRLVYGLTK-----LTAGT 996
QY 1345 ETFFSWTYYSKQEGNELDKELKEGKIPPAFLRSMFTYTGDRDLFGDISKHGEGS 1404
Db 997 SS-----ESPQG-GESESR-----ASDVSOQNG-- 1018
QY 1405 KLKEQJDSLEFKNGDQKSPNKTRQEWMTESHSHEIWEAMLCAIVKIGAKKDFENYGN 1464
Db 1019 -----1018
QY 1465 VKESDSTLEBAPKRPQFLRMLTEWYDYCYTROKYLKDOECKSNOLKCDTECNK 1524
Db 1019 ---GDITTTESLRK-----WFE-----TAA 1037
QY 1525 CEDYVAKMKKKWEIIPDKTYKDERKKRPDRQIHGVMTDYGTNATDYLNKRTASCG 1584
Db 1038 IETFFLMHRYKKWEAOKR-----1056
QY 1585 DKFGSASYORNIQLEKQAYYADAKHGCTKFIENDDKYTINISSKDKGLVEANTGA 1644
Db 1057 -----AELORNGLL-----GTGA 1070
QY 1645 IKWONKGPNNYNNIKELTEDVLEPSSRLRICFHALDGNYPDEYK--DENGIRKRLME 1700
Db 1071 S--LNLGGDSNPOTOLQKSGTIPDLRLMEFTL--GDYDILVRYGVAADKNG-----1120
QY 1701 VAATEGNLGOYYKKEKKEKIKTSDAHKYSYEVPPCSAMKYSYDLDIILGIDNLEDE 1760
Db 1121 ---GNN-----TILNASGNKE 1134
QY 1761 KORTE---ENLKIIFNKNGTSVGKSDSTGNPGSTARKEFFMNEKNENANMIG--YK 1815
Db 1135 KOKMEKIQEKIETOLPITSGKKEKTRGPONSVD-----KSLMDIRIAHWHVHGWYCALTYK 1189

QY 1816 RGRDDNSGNSANSDEDLKCGSVSPDDDYPMGRNDEGTAYOLFMAEMGEDFCNKE 1875
Db 1190 ---DDONG-----1194
QY 1876 KELBKLVACNDYTCGDNEDKRRKCTDACYKKFISEMKPOYEKOIKKYGENKDIYSE 1935
Db 1195 -----1194
QY 1936 HPAKOADEAREYLDQOLKIKCENKSGDCYKCMKQVSTQRLDNGSONNPASLDEPERE 1995
Db 1195 -----LKV-----1198
QY 1996 VEGKCNQVRCGPVRVRETPSPRVLISKATASKKEAKTAPPTQPKVENUTTEBRAQ 2055
Db 1199 -----VKKQKXEN-----1207
QY 2056 TTRRRAAQTRKRSTATTSTESDVGTWVKAILSKPDSRGIEGKPNKTYGQIPKMGCIY 2115
Db 1208 -----1207
QY 2116 GSKENENGICMPRRKRLCINNIOYLVETENKRONDIKAFIKCAIEFQFLMYII 2175
Db 1208 -----PEK-----LM-----1212
QY 2176 ENPAENELONGTIPDEKFRIMYYTGDKRMEFTDISDKKITVNSVTLNENNK 2235
Db 1213 -----NETTK 1217
QY 2236 KODKKADEELRKIFWENKKNKFIWEGMIGLTHLTDENEKERIRDNYQINDMTKTLPSL 2295
Db 1218 KPKDEKYOYQAKL-----EDESERGEKRPDSASG--TKLT-- 1250
QY 2296 EEFYKRPOLRWPEMAEFCNKRKEOLKLEAGCKRYEONGSDGTQCAE-----2348
Db 1251 -DFTKRPYRYLEWGENFCRKTETMGIKEDC--YKNGRCSGGLCNEIYVIDEK 1307
QY 2349 -----ACVYONFIRKKKTEYEROREKFKDKDKKKYKDYSTERDIKATCAHEYL 2400
Db 1308 IFGDLPCPCARICRFYKKW-----INTKRD-----1333
QY 2401 NMKLELCGNKDCSMOKPSSOLPKTTQOSSANDMPSLDYVPEFPNKCCEBELSK 2460
Db 1334 -----EFNK-----1337
QY 2461 GSMHTKITEPKIPMNCVEKAAYLSKEAENNDITLKEKFIPIESTKESKNSFTNN 2520
Db 1338 -----OSNAYSQKKRY-----BEENDSAOKNN 1360
QY 2521 NPCDPKRPYAPDKYGR--RNPCENRENEKYDYEMKCYKNSKFYOEKRVCPVPRRH 2578
Db 1361 GVGCTLKDDAAE--FLNLTAKNGPCNKESENNKKADELDFPKPBDTFFDADN--CKP-----1413
QY 2579 MCLRLNDEIKIERKDSNYLLAKMYRTARNEGIDIIKFNFSNCGCANPICD--TMKYSFA 2637
Db 1414 -----CSEFKIK-----CENHSSCGSNTQGRCDGTTTIAAT 1445
QY 2638 DLGDIYRGV--DMLRIGYLPVEIKLYKVEYIYKMRKKNKGRKRYNDYQFRSAAW 2694
Db 1446 EIEIKKTNTKEVYMLVADDSKATEFK-----DGLSECKDKGIFKGI-- 1487
QY 2695 DANRKDIWKAMTCAPADAKLFRKGRMDGERITL--IQDKCGHMDPVPDYIIPQPRMM 2753
Db 1488 ---RKDEMEC-----GKYCGYDICMLKKKNDIGRESD--KKYIIMK--ELL 1526
QY 2754 TEMSEYCKALMELEKFKSCHCKTSIDCKKNDYDNKCEQCTROEYKKNYPLAKKSL 2813
Db 1527 KRWLETF--LEDYNNIKIKHISHTKNGK--GSKC-----IKGCVDKW-- 1564
QY 2814 FDIOSNRYKELYEQPIYTKISTYDVONFQKLTFTKSECSVESSEYELHETSKCLYKF 2873
Db 1565 -----VOOKKEEMKQIK--EFNNE-----QYKS 1585

QY 442 DFFLNLNBECKYCKG---GLPBEKDTFTNSADDKGIFRSEYCOVCPCCYKCDGICY 497
Db 461 DKFLKILNKSGIOKOPQOVENKADNVDFNEKYVK--TFSTRICEPCPCPGCEKGGPPW 519
QY 498 THESDNDREVNNDYPRPGVAKPTNTVLVSGNEODDIQOKLENPCNSSTNYKDKNNOK 557
Db 520 KYVGDKTCGSAKRTYDPK---NITDIPVLPKPSOONILKATKNCCEKAP--GGGOIKR 575
QY 558 WECYKD---ENINCKLEONTEINNDPKIISFHEFELWVYLLRDTIKMNDKLT 612
Db 576 WOCYDEHRRSSKNNNKCTKCCGCFORKEVKOEEMAIKJHFKQDIDYQ---KGLI 690
QY 613 CINTTT---HCIDECNRNCLCPDRVAKOKEEMNSIKLFTKKNKIOOSYISNINLP 668
Db 636 CINNNTNGNTCRNNKCKTCCGCFORKEVKOEEMAIKJHFKQDIDYQ---KGLI 690
QY 669 EGYEFYMDKLDDEAKMKELMENIKRKNFNSLENNRYLE-----NAI 714
Db 691 VEPFYGLVLY---LKGNLONIKDVGDTDDIKIKILDBEDAVAVVLOGKNTT1 746
QY 715 ELLDLKETATICKDNTNEACETSHNATTNCPVCRGTOPTKNIKEIAQYFKRSAYE 774
Db 747 DKLOHKEBAEQCKK--QEBC-----KAAQ 773
QY 775 EARNRGLHLKGAHAGIYKRGRRKDFKDNLCRIMKHSNRLGFSNGPCDCKGTGDI 834
Db 774 ESRGRSAE---TREDER-----TQPADSAGE--- 797
QY 835 QTRFVGTETWEVDPEHMRKHEDVIMPRRRHICTSNLEHLOTDDHNLNINVDLYNNS 894
Db 798 ---VEEEDDDYDEDEDE- ---DDVVQEE 820
QY 895 FLGDVLLSAKYEANKIIRMYKERNLKGPEVTDPKHOTTICAIRYSFADIGDIINGRD 954
Db 821 EEG---KEEGTVEVTEVEVEET----- 842
QY 955 LMERNGDMVLOGLHETVFNINIKSLKKGKNDXYNDAPRYLLKRENMWEANRKYWEAM 1014
Db 843 --- 842
QY 1015 KCDIKYLDKSGHSTQSYCGYSDHTPLDDYIPQKLRMTWMAEWYCKVQKEYDLKE 1074
Db 843 ---VTE-----QE 847
QY 1075 KCECKDKDNGOGCTRESGTCTKTEACNEYNDIGLMEQWNIISDKYKELHEQAOMS 1134
Db 848 GVRPC---DIYVKL----- 858
QY 1135 VNSNGIEASSTAKNHIDRNVIEFLSELVQONGKSNKSGTSDSAVIGTWTYENGCAYL 1194
Db 859 --- 858
QY 1195 HDTGNDPDCOSQNFCDKSDKDNKRYAFRDKPRQDHGACGCKSGKPTRVQIKTKKKA 1254
Db 859 ---FEDDKSLKE---ACGLKYG- 875
QY 1255 EERKOTBEKTYNDILKENDGKROYEDCHPRKNSNGYRPMOCGINLV-----EDPRYCM 1308
Db 876 ---GGEK-----FRMKCVTPSGVSTANSGDGAICVP 906
QY 1309 PRROKLCVHFLA---NDNEIKKLOSQV-----NLKEAFIKSAAEFFSWYY 1353
Db 907 PRRRRLYVGLSOWASRGSEJTEVSSEASAPSOSESEKLRATFISAAIEFFELHKKY 966
QY 1354 K---SKDG---EGNELDKELKE--GKIPAPLRSFVYFEGDVRDLFLG 1393
Db 967 KEKKKPRATDGAAGLVSLEPSPRGEDPOTLOQOTGVIPFDLRLQMFYLLADYKDL- 1024
QY 1394 TDISKHGESKLEQIDSLFKNGDKSPNGKTRQEWTEHSHIWEAMCALVYKIGAKK 1453
Db 1025 --- 1024

QY 1454 DDFENTGYNNVAFSDKSTTLEEFAPKRPQFLRMLEWYDDYCTYRQKYLADVOEKCKSND 1513
Db 1025 --- 1024
QY 1514 QLKODTECNKCEDYVYKMKKKKEMIPQDYKYEDERDKRRFRQHGIGVMTDYGTNATD 1573
Db 1025 ---YSGNDTS 1032
QY 1574 YLNRKFTASCGDPRGSASVVOIRNIOLEKQAYYDADKHCCTKEIENDRYTNISKDKC 1633
Db 1033 ---DTTGQTPSSNDNL 1047
QY 1634 KGLYKEANTGAIKWQNGPNYNNLKLTELDELVLFPSRLICFHALDGNATDEPVKENG 1693
Db 1048 KNYLEAS--- 1055
QY 1694 LRRKIMEVATEGYNLQYKKEKKEKIKTSDAHKYSYEVPCASAKYSFYDLRDIILG 1753
Db 1056 ---GST- ---OEKKK- 1066
QY 1754 IDNLEDEKOKTEENLKRIKFNKNGTSYVGKSDSTGNPSTARKFEMNKECYWAMIGC 1813
Db 1067 ---QIOAKIKKILN-GATS---GVPPYTKNSVKTPOQTMWENIKADITMANAVCA 1113
QY 1814 --YKRGDNGSGNSARSDEDLKCKGSPSDDDYPMGKNRDEGTAYQFLRPAEWGDFC 1871
Db 1114 LTYKE---NDAR-----GTS----- 1125
QY 1872 KHKEKELKLVGACNDYTCGDNEDKRRKCTDAGCYKKFISEMKPOYKQIKYGENKDK 1931
Db 1126 ---AKIOKND- 1133
QY 1932 IYSEHPYAKDAEAREYLDOLKRIKICENSGDCEYKCMKDVSTORTLQNSQNPASILD 1991
Db 1134 ---LKK-----ALMDEANKNP- 1147
QY 1992 EPKEVEGKCMQVPRGPRVRRRETSPRVSLSKATSKKEAKTAPPTOKPKVENLTTE 2051
Db 1148 --- 1147
QY 2052 MRAQTRTAAQOTRRKRTSTATTESDVGTMVKAILSNKPRDSRGIEGCPKTYGYPKW 2111
Db 1148 --- 1147
QY 2112 GCIVKSKENENGICMPRRRKILCINNIOYLANETENKRNNDIKAEAFKCAIETOPMLT 2171
Db 1148 ---IEKYQYTNVLE----- 1159
QY 2172 KYIIEPAENELONGTIPDEFKRIMYTYGDIKDMFFGDISNDKIIITVNSVTIILN 2231
Db 1160 ---D 1160
QY 2232 ENKKKQDKKDELRKIFMEKNNKFIMEGMIYGLVHLNDENKERRIDNYQYNDMTKL 2291
Db 1161 ESKAKSND---TIQ 1171
QY 2292 TPSLEEFKRPQFLRMTEWAEFCNRRKEDOLKLEAGC-----KEYEONGSNDGR--- 2342
Db 1172 PPTLKNFVEIPTFRMLHENGNSCFERARAKIQAIRKECHDEDOEKY---SGDGYCE 1227
QY 2343 ---TOEBAKCVIYQNFITKKTETEBEQREKFKDKQKQKYYKDYPTSPERD 2389
Db 1228 EIFSROYNVLODLSSCAKCRLTKYTIKKEKTEYERQOKAYEQOK---SNYENQKD 1282
QY 2390 IERKATCAHEYLNNKLKELCGNKDCSQKQPSQLPKTTQOSQSSDADMPESLDYVP--E 2447
Db 1283 ---KC-----OTOSNNANEFKRTIGASPTAA 1306
QY 2448 EFNKCEPELSKSGSMITHRTKITEPKIPMNCVERAAAYLSKEAENNDITLKEKFIPIES 2507
Db 1307 EF---LQKLS---CKNDNGY-----ENGEDNKIDFR-NPDKT 1337
QY 2508 TKESKSNMTNNNPNCPDKPKYAPDKYIGRRNCPENNEENRKFVYDMKCYKNSKPYQEK 2567

```

Db 1338 FKEAHS-----CDP-CPTGVK-----CQNGH-----1358
QY 2568 KRVCVPRRREHMLRNIDELIKIERLSNYLKKMVRTARNEGIDIKNFNSGCMANP 2627
Db 1359 ---CVGSANGKECKNN---KITAEDIKN-----KTDNGNINEMVSDSTN-----1398
QY 2628 ICDMKYSFADLDIYGTMLRIGYLYLPEVEIKLYKVFYIYGKWRNKNKGRNRYNDVQ 2687
Db 1399 -----TFEHLGD-----CKSSGIFKGI-----1415
QY 2688 TFRSAMWADRKRDIWK-AMTCAKPEDAKLFRKRGMGFEITLIOD-KCHKDDPPVDY 2745
Db 1416 -----KDEMKCANVC-----GVDICTLEKIKIKNGQED---KKY 1447
QY 2746 IPRFRMTWMSYCYCCALMELEKFRKSCDHC-KTSDRCNDYDENKCEQCKTRCOEYK 2804
Db 1448 ITWK-ELLKRWLEFY-----LEDYNRIKKIKLCTKKEDGK-----CIKGIIE- 1490
QY 2805 NFVKKMSLFIDIOSNNKYKEIYEPITTKISTYHVQNFVQKLTKEKSECSVESFSEYLHE 2864
Db 1491 ---KM-----VO-----EKTKEWO-----1501
QY 2865 TSCKLYKFNENDSSNIRTYAFETPKSYKACSCSLPSKPNLDNCPTDQNDGCKEIQ 2924
Db 1502 ---KIND-----TY-----LE 1509
QY 2925 TTFPCSKNDYDNLDMMNAYLVNNSDDKGVILIPRRRRLCTRPITAYNRYRKDKEIYL 2984
Db 1510 QY---KNDGNTLTF-----LEQOYR-----1529
QY 2985 KKLITSAFSOGQLLGOKYKSEELCFEAMKYSYADYSDIIGKIDMDMDSLSKIKKIET 3044
Db 1530 ---TEFKNAIKPCDGLD-----1543
QY 3045 SNEATENKRTWMENNRQIWMHMLGKYIATSKVTLDEGWCQLPXDENQFLRWLIEMA 3104
Db 1544 ---QF-----1545
QY 3105 KOACKERKHVSDSLKTCPRSNEDNFEASELLRQPCCONDIRKYSIATLNTNMENTNI 3164
Db 1546 ---KTSCLNSTDN-----SQNGNNND-----LYVCLLNKLOK 1575
QY 3165 KTKQKLDKSSGNIDNKPSEENVQYIKSDSOCALELDINEIYVTKNNENN-----3217
Db 1576 KISECKEQHSGQTOR-PCD---NSSLSGKESHLVEDVDYEE---QNEPKVQOPKFC 1626
QY 3218 -FEKEVAKLYLGLYFVEDETHKNVLDGNT--KEEBQTVRPKALYFTPHVDSFYQAPL 3274
Db 1627 PUMKEPKKENDVEVGTGCGDEEKKEVDSVLEQKEEEMASAPESPPPLTP-----EAPK 1680
QY 3275 FSTHVAQYDP-----KN-----DILKGSISVIVSALGL-----IALHFMKKRFRS 3316
Db 1681 KEENVVPRPPPRKRRRIKTRNVLDHPAVITPALMSSTIMWSIGIFAAFTFYFLKKRKS 1740
QY 3317 SV-DLLRLILNIPQEGYGMPTLESKNRYIPYRSGPYKGTIYMEGDTSGDEKYMWDLS 3375
Db 1741 SVGNLFOLIOIPKSDYDIPPLTKSSNRKIPIYASDRHKGTIYMEGDSGDE-KAFKMDT 1799
QY 3376 SDITSESEYEELDINDIYVGSPPYKTLIEVLEPSKRDIPSD-----DTGSNDTPR 3428
Db 1800 TDITSESEYEELDINDIYVGSPPYKTLIEVLEPSKRDIPSD-----DTGSNDTPR 1859
QY 3429 TNRFLDDEMNLKHDVFSQYL---PNTPEPNNNYKADIPMNTPEPTLYSDNPEEPFIS 3485
Db 1860 P--ITJDEWMQLOKKDFISNMLONTQNTPEPILHDNV--ANTHP-TMSRHMDDOKPFIWS 1914
QY 3486 IHDRLYIGKELSY-----NINNSTN-----TNDIPMANARNSRYGIDLINDS 3529
Db 1915 IHDRLYIGKELSY-----NINNSTN-----TNDIPMANARNSRYGIDLINDS 3529
QY 3530 LVVNLNLTLY---MMK 3541

```

```

Db 1975 LSGNHIDIDYDEMLK 1988
RESULT 10
AA77906
ID AA77906 standard; Protein; 2182 AA.
AC AA77906;
XX
XX 13-JUN-2000 (first entry)
XX
XX Plasmodium var-1 polypeptide.
DE
XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Stialic Acid Binding Protein; SABB; malaria; vaccine; immunisation;
KW protozoacide; var-1.
XX
XX Plasmodium sp.
OS
XX US593827-A.
PN
XX 30-NOV-1999.
XX
XX 07-JUN-1995; 95US-0487826.
XX
XX 10-SEP-1993; 93US-0119677.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
XX WPI: 2000-194198/17.
XX DR N-PSDB; AAZ98288.
XX
XX Isolated protein binding domains from plasmodium vivax and plasmodium
XX falciparum erythrocyte binding proteins useful for vaccinating against
XX malaria.
XX
XX Disclosure; Columns 129-140; 93pp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
XX Binding Protein (SABB), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABB are the
XX respective ligands for plasmodium vivax and plasmodium falciparum Duffy
XX and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX Immunization with the polypeptide provides effective protection against
XX malaria. The present sequence represents the var-1 polypeptide.
XX
XX Sequence 2182 AA:
SQ
Query Match 9.0%; Score 1743; DB 21; Length 2182;
Best Local Similarity 20.3%; Pred. No. 1,9e-95;
Matches 757; Conservative 339; Mismatches 678; Indels 1960; Gaps 126;
QY 13 GNAASLEGDAKSPILKESHS--ARNVLER-----YAKNIRHPSYAKEHVDLSKGLDK 66
Db 10 GSGGSS-SGKGGKDTSEYIYVSAKDLDDVGEKYEKKNKD--AKKTEALKGNLNT 66
QY 67 AEPFGSGSTYVKNHNYPPPCNLDHKEHTNLRYDVN---LRPHG---GREQNRPE 119
Db 67 ANGSSSETASSIE-----TCPL-VKEY-----YERYNGDGKRHPCKRDKAKNEDVNRFS 114
QY 120 DESEEG-NKIRNYKRRKNDI-ACAPRRRHMCDKMLELNDINTQNIHDLGNVLYTK 177
Db 115 TLGGQCTYNAIKKSOQDNKVGACAPRRRLHCLDYNESED--TSTTKRLLEVCMAAK 172
QY 178 YEGESIVNNH--PHKGT-----SDACTALARSFADIGDIYRGIDMF-----KPNVHDV 224
Db 173 YEGNSI-NTHYTOHRTNEDSASQCLCTYLARSFADIGDIYRGIDLYLGIDNKEKEQRK 231

```

0Y	225	ETG:REYFPRKTHDMEDEVND-----YNPDG--SGUYUYLREAMVNNKRNKWEJLITCDA	278
Db	232	BOKLADIPKFKH---KVMKTNGAOERT IDDAKGDFQFOLREDMWMTSNBEYWKALICHA	288
0Y	279	SYKSGYFMOSESNTPLFSNPRCGHOGKQVPTNIDYVQYUPLRMFDEMEGEFCRRKRIKTK	338
Db	289	PREANVFYKTAACNAGKGTNOGCHIGSDVPTDYVQYUPLRMFDEMEGEFCRRKRIKTK	348
0Y	339	VKDSGRNDKEBLYCSHANGHDCITTTIMKGLIHLDMKCTDOSTCKVCEYEWLGNDOEAPRK	398
Db	349	LQKOCRDYEOUMLYCSGNGYDCTKIYKAKGLYGEHCJTCNSWVCBRYEWTIDNOKEFLK	408
0Y	399	QKEYEKEI-----QSYLSDNRKRVNNINSRYUYQYFEYKLEKEYOYATN	441
Db	409	QKRIETEISGCGSGKSPKRTKRAANSSSSDDN-----GYESKFFKKLEGEYDOV	460
0Y	442	DTFLATLNEGKYCGK---GLPGEKIDITTFNSADDKGIFYRESEYCVPCDGCVKDGIKY	497
Db	461	DKFLKILNKBEITCQKQYOVENEKADANDVFNEXUYV--TTSRTEIGPCWCGLKEGGPPW	519
0Y	498	THKSDNDRERVANNEDYPRPMGVKPTNITVLYCSNEGODITQKLENCNSSTNYKDKNNOK	557
Db	520	KVKGKGTGCSAKTGYTPK---NITDIPVLYPRKSOONILKTKYKNCEKAGP--GGGQJIK	575
0Y	558	WECYKID-----ENINRCLEOMTEINNDNPKIISPHNFELWYTYLLDITKMNDKIT	612
Db	576	WQCYIDENRBSKNNNNCCVEGTWDFKFOGQYKASTVNEFFWDMVHLDHVSXKTELSK	635
0Y	613	CINNNTT---HCIDECNRNCLCFDRWVKOKEEEMWSIKKLEFKKKNOOYSYNNMLF	668
Db	636	CINNNTNGTNRNNNKKCTDCCGCFQKWEKKOEEMALIKDHFGKQIDYOQ---KGLI	690
0Y	669	EGYFXYVMDKIDKDEAKMKELMEIKKKKEBNLENNDYLE-----NAI	714
Db	691	VSPSPGVLDV---LKGMLNLOTKOVHGDJTDIKNIKKLDEEDBAVAVLGKDNNTI	746
0Y	715	ELLDDHLETKETICKDNNTNEACESTSHNATNPGVPRGOTPTKNIKEIAOYFRSAVE	774
Db	747	DKLLOHEKQNEGCKOK--QEBC-----KKMOQ	773
0Y	775	EARNRGLHKLKGAHBEGLYRGRGRDKDNLCRIMIKHSNNRNLGFSNPGCDGKGTGDI	834
Db	774	ESRGRSAE-----TREDER-----TOOPADSAGE---	797
0Y	835	QTRFVGVGEVMEVDEHMRKNDHEVIMPPRRRHICTSNLEHLOTDHPRLGNIVDVLNNS	894
Db	798	-----VEEEDDDDYDEDED-----DDVYEE	820
0Y	895	FLGDVLLSAKYEANKIRIYRKKEKNLKGPEVYTDPKHOTTCIRAIYSPADIGDIIIRGD	954
Db	821	BEG-----KEBOSTVIEVEVEVEBT-----	842
0Y	955	LMERNGDVAKLOQHLETVFGNIHKSJLKGKNDXYNDAPRYLKLRENMEANRAKVEEM	1014
Db	843	-----	842
0Y	1015	KODIKYLYKDSGHQSTQSGYSGYSDHTPLDDYIPQKLRMTTEAEMWYCAVQKKEVDKLKE	1074
Db	843	-----VTE-----OE	847
0Y	1075	KCKECKDKDNGOGCTKESGTGCTKTEACNEYNDIIGLMEOWNNISDRYKELHROAOMS	1134
Db	848	GVPKPC-----DIVGKL-----	858
0Y	1135	VSNSGIEASIAKHHIDRNVIEFLSELVQONGSKSNKSGTSDSANVIGINTYENYAVL	1194
Db	859	-----	858
0Y	1195	HDYGNFDDCQSONFCDCEKSDKDNKEKYAFRDKPODHGACGCKSGSKPTVQIOTWKKA	1254
Db	859	-----FEDDKLKE-----ACGLKTYP-----	875

QY	1255	EKKTEKTVNDLTKENDGKKOYEDCHPPKNSNGYDMDGCGNINLY----	EDPRVCMP	1308
Dd	876	-----GGEK-----	-----PNNMCVTSPGCVSTATSGKDALICVP	906
QY	1309	PRROKLCVHFLA-----NDNEIKIKLOSQV-----	NLEKAEIKSAAEFEFSWYU	1353
Dd	907	PRRRRLYVGLISQWASRGSGDETTVESSSEATSAPOSSESEKLTARFIESAALETFFLMHY	966	
QY	1354	K-----SKDG-----	EGNELDKEIKE-GKTRPAFLASMYTIGODRDLFG	1393
Dd	967	KEEKKPRATQAGAGLYSLPEPSPPEDDPTQLOQGVLPBPLQMEFTYLAADYDLI--	1024	
QY	1394	TDISKHGEGESKLEKEQJDSLFFKNQDGKPSGKTRQEMTTEHSHEIWEAMLALVKGAKK	1453	
Dd	1025	-----	-----	1024
QY	1454	DDFEENTGYNNVKEPSDKSTLLEBPAPKPOFLRWLTWYDICYTROKYIKLDQVECKSND	1513	
Dd	1025	-----	-----	1024
QY	1514	OLKODTECNKKCEDYVYKMKKKKKEMIPDKYKYKDEBKKRPDRQHIQWATDYDTGNATD	1573	
Dd	1025	-----	-----YSSSNTPS	1032
QY	1574	YLNKFTASCGDKPGSASVYQRIOLLEKOAYYADADKHGCTKFTENDKRYTNISSKDKC	1633	
Dd	1033	-----	-----DTGKQTPSSSNDL	1047
QY	1634	KGIYKEANTGAIKWQNGPNYNNIKELTPEVLPFRRLRIGCFHALDGYTDEPKVDENG	1633	
Dd	1048	KNIVLEAS-----	-----	1055
QY	1694	LKRILMEVAATEGYNLGOYYKEREKEREKIKTSDAHKSYEVPSCAMKYSFYDLIDILG	1753	
Dd	1056	-----GSTE-----	-----QEKKKK-----	1066
QY	1754	IDNLEDEKOKTEENLAKTIFKNGTSTVGSGSDSTTGPGSTAKFRMNEKKECVNMAATCG	1813	
Dd	1067	-----QIOAKIKKILN-GATS-----	-----GYPPVTKASVTPOOTWEMENIANKIINWAMCA	1133
QY	1814	-----YKRGDDGNSGASASDEDLKCKGSPSDDDYPMGRNRDEGTAYOFLRMEFEMGEDFC	1871	
Dd	1114	LYTKE-----NDAR-----	-----GTS-----	1125
QY	1872	KHKKELEKLVGACNDYTCGDNDNRKRRKCTACTOYKRFISEMKPOYERKOIKRYGENDK	1931	
Dd	1126	-----	-----AKIQOND-----	1133
QY	1932	IYSEHPAKDADEAREYLDKOLKIKIGENKSQDEYKCMKDYSTORULTGNSQNMNASLSD	1991	
Dd	1134	-----LKK-----	-----ALMDANKNT-----	1147
QY	1992	EPKEVEGKCNQVPRGPPRYVARETPEPBRVSLISKATASHKEAKTAPTRKOPKEVENTTE	2051	
Dd	1148	-----	-----	1147
QY	2052	MRAOTRPRRAAOQTRKKTSTATTTESDVGTMVAALISNKPDRSGIGBCNPRTTYQOYPM	2111	
Dd	1148	-----	-----	1147
QY	2112	GCIYGSKENENGICMPPRRKKLCINNIQYLANETENKRRNDIKEAFICAIFOTPLML	2171	
Dd	1148	-----	-----IEKYQTVNKL-----	1159
QY	2172	KYIIEENPAENELONGITIPDEFKRIIMITYYGYKDMFEGTDISNDKKIITYVNSVTTILN	2231	
Dd	1160	-----	-----D-----	1160
QY	2232	ENNNKKODKKDEBLARKIIEWKNNKKFIWEGMIYGLTYHLTDENEKIKIDNTQYNDMTKL	2291	
Dd	1161	ESGAKSND-----	-----TIQ	1171
QY	2292	TPSLEEFKRPQFLRWTFTEMAEEFCNKRKEQILLKEAGC-----	KEYECNGSNDK--	2342

```

Db 1172 PPTLKNVETPTFRWHEWNSFCFERAKRLAQIKHECMDEDEKQY-----SGDGEYCE 1227
QY 2343 -----TOEACAVCYONFLKKKTEYEROREKFKDKGKTKYKQYPSIERD 2389
Db 1228 EIFSKQYNVLADLSSCAKCRKLYTKTEKKTEYEROKAYEQQ-----SNENNOCKD 1282
QY 2390 IEKATCAHEVYIMKLKELCGNKDCSCMOKPSOLPKTTOOSOSSDANDMESLIDYVP--E 2447
Db 1283 -----KC-----QOSNNNAEFSGTLGASPTA 1306
QY 2448 EFNKCEPELSKSGSMHTTKITPKIPMCVEKAAYLSKEANNMDITLKEKPIES 2507
Db 1307 EF-----LOKLS-----CKNDNGY-----ENGDNKIDRK--NPDKT 1337
QY 2508 TREKESKNSWTNNNPPCDPKPYADKYIGRRNCPENRENERFKVDYEMKCYKNSKFQEK 2567
Db 1338 FKEAHS-----COP--CPITGVK-----CONCH----- 1358
QY 2568 KRYCVPRRREHMCRLNIDETKIERLKDSTNYLKMVRTANEGIDIIKNPSENCGAMP 2627
Db 1359 ---CVGSANKECKNN--KITAEDIKN-----KTDPNGNIEMVSDSTN----- 1398
QY 2628 ICDTMKYSFADLDIYGTMLRIGGLPVEIKIKVFEYIYGKMRNKKGNKKNNDVO 2687
Db 1399 -----TFEHLGD-----CKSSGIEFGI----- 1415
QY 2688 TFRSAMWANDNRKDIWK--AMTCKAPBEDAKLFRKGRMDGFERITLIOD--KCGHKDDPPVDY 2745
Db 1416 -----RKDEMKCANVC-----GVDICILEKKIKKQBED---KIT 1447
QY 2746 IPORFWMTEMSRYCKALMEELERKKSODHC--KTSDRCKANDYENKCEQCTRCQOEYK 2804
Db 1448 ITMK--ELKRWMEYF--LEDYNRIRKIKLCTKKEDECK-----CIKGCEI-- 1490
QY 2805 NEVLAKWSLFDIOSNKRYKELYEQITYKISTYDVONFQOKLTFKSECSVSSEFSYLYE 2864
Db 1491 ---KW-----VQ-----EKREMO----- 1501
QY 2865 TSKCLMKNENDGSSNITYAEETPKSYKACSTLPKSNPLDNCPTDNKCKEQLQ 2924
Db 1502 -----KIND-----TY-----LE 1509
QY 2925 TPTFCSKNDYDNNDNMNAYLYLNSSDNKGVLIPRRRHLCRPTITAYNRKKGDEIKL 2984
Db 1510 QY---KNDDGNTLTF-----LEQOYR----- 1529
QY 2985 KRLITSASOGOLLOQYKSEBELCEFAKMSYADYSDIITKGTMDMDSLSEKIKKIFET 3044
Db 1530 -----TEFKNAIKPCDGLD----- 1543
QY 3045 SNEATENRKTWENNRRQIWMHMLGKYKATSKYTLDEBQWCLPDEETNOFLWLEMA 3104
Db 1544 -----OF----- 1545
QY 3105 KOACEKKHVSLSLTKGPRSNEDNFEASELLROPCONDIRKYSILNLIKNTMENLNI 3164
Db 1546 -----KTSGLNSTDN-----SQGNND-----LYCLINKLQK 1575
QY 3165 KYKOLKDOSSGINDKPSBENVOSTIKSDSCALELNDIMEYGTNNENN----- 3217
Db 1576 KISEKEHSHSGOTQ--PCD--NSLSCKESTYLVDDVDEE--ONPKKVVQPKFC 1626
QY 3218 -EFKVLKLYPGLYFVDETHKNHYLDGNT--KEEDQTVPRKALYFTPHVDSYQAPL 3274
Db 1627 PDMKPKKENDEEVGTCGGDEKKKVEDSYLEQKEEASAAPESPLTP--EAPK 1680
QY 3275 FSTRHVAQYD-----KN-----DILKSSISVIVLSAGL-----IALHFKMKKPKFS 3316
Db 1681 KEENVVPRPPPPKRRIKRTNVLDPHAVIPALMSSTIMSGIGGFAFYFLKKTKS 1740
QY 3317 SV-DLRLINLPOGEYGMPTLESKNRYIPYSGPKKGTIYIMGDSGDEKYMMDLSS 3375

```

```

Db 1741 SVGNLFOIILIPKSDYDIPFLKSNRYIPYASDRHKGKTYIYEGDSSGDE- KYAEWSDT 1799
QY 3376 SDITSSSESEYEELDINDIYVPGSPKRYKTLIEVLEPSKRDISD-----DTPSNDPR 3428
Db 1800 TDIITSESEYEELDINDIYVPGSPKRYKTLIEVLEPSKRDISDINDIIPSDINSDTPR 1859
QY 3429 TNRFIDENNELKHDVSOYL---PTEPRNNKSKADIMANTEPNTLYSDNPEKPFIT 3485
Db 1860 P--ITDDEMOKLKDFTISNMLQNTQTEPNILHDND--NTHP--TMSHHNDOKRPFIS 1914
QY 3486 IHDRDLYTGKELSY-----NINMSTN-----TNNDIPMARNDSTYRGIDLINDS 3529
Db 1915 IHDRNLFSGEYNYDMFNGSNPNINISDSTNGMDSLTNNHSPYNDKNDLYSGIDLINDA 1974
QY 3530 LVLNLLIT--MAK 3541
Db 1975 LSGNHIDYDEMILK 1988

RESULT 11
AAM93944
AAM93944 standard; protein; 2228 AA.
AAM93944;
30-JUN-1999 (first entry)
P. falciparum PfEMP1 protein.
Erythrocyte membrane protein; EMP; PfEMP1; malaria; antiloocclusion;
glycosamino-glycan-like moiety; antiaggregational; antimalarial;
antigen receptor; infected erythrocyte; rosette formation; blood cell;
capillary occlusion; cerebral malaria; treatment; vaccine; detection;
medicament; parasite; diagnosis; drug screening.
Plasmodium falciparum.
W09915557-A1.
01-APR-1999.
18-SEP-1998; 98WO-SE01675.
19-SEP-1997; 97SE-0003386.
(KARO-) KAROLINSKA INNOVATIONS AB.
Barragan A, Carlson J, Fernandez V, Qijun C, Wahlgren M;
WPI; 1999-254692/21.
New isolated malaria polypeptides
Claim 4; Page 67-74; 80pp; English.
This invention describes a novel Plasmodium falciparum erythrocyte
membrane protein (EMP), PfEMP1, which is capable of binding to a
carbohydrate which exhibits at least one negatively charged
glycosamino-glycan (GAG)-like moiety and has antiaggregational,
antiloocclusion and antimalarial activity. The carbohydrates of the
invention are capable of acting as receptors for malaria antigens
present on the surfaces of malaria infected erythrocytes, by binding
to these antigens the carbohydrates prevent erythrocyte formation by the
blood cells, this prevents occlusion of capillaries as is seen in
cerebral malaria caused by Plasmodium falciparum. The products of the
invention can be used to treat malaria or to vaccinate against it, or
used to design a model to identify compounds that bind to PfEMP1. The
carbohydrates, polypeptides and antibodies of the invention can be used
as a medicament for dissolving the rosettes formed by erythrocytes
infected by a malaria parasite. The products can also be used for
detection, diagnosis and drug screening.
Sequence 2228 AA:

```

Query Match 8.7%; Score 1601.5; DB 20; Length 2228;
Best Local Similarity 20.1%; Pred. No. 9.9e-92;
Matches 754; Conservative 330; Mismatches 673; Indels 1985; Gaps 137;

33 KSNRVLEERAKNRRHSK- -AKEHDSLKGLTAEFFGSGPSTPVNKNINYYPPYCNL 90
13 EDKHNVLDERGQKV- -HDEVLGEAKNYSLEKGSLSLSTLIGETAFYVK- - - - - 59
91 DHEHNTLRYADV- - -NLHHP- - -HGRENRFDEDESECGNKIRYKRN- -NDAIAC 141
60 --SMQFSKSTELLEANSKRNPKCKKQKQNDVRFYKBEAGD- - -KMCSSNATC 113
142 APPRRHCKNLEALNDITONT- -HDLGNVLTAKYEGESIVNNHP- - -HKGS-D 194
114 APRRLHLCKNKNPNMNSNDSKAKHDLAEVCAAKYEGESIKTHPKYDSKPGSDFP 173
195 ACTALRSFADIDIGDIVGIMF- - -KPNVDKXETGLREVEFKIHGMD- -EV 243
174 MCTMLASFADIDIGRLDYLGNKKKQNGKETEREKLEOKLETKETKIHDLKQKEA 233
244 KNDYNPDSGSGNYKLRAMNNVNRNKNWEAITCDASY- -KSGYFMQSESN- - -PLFSPNK 299
234 OKRYNGEDBNFYKRLRDMWTANREIYWGAMTSCKELDNSYFRATCNDTGQSGPSTHNK 293
300 C- - -GHQGR- - -VPTNLDYVPOYLRMFEDWGEBCRNKIKLKVKVDC 343
294 CRCDKKGANAGAPKADGDVTVPTFYFDVPOYLRMFEMADDFCKKCKKLEMLEKOC 353
344 R- -NDR- - -ERYLCSHNGDCTTITMKGILHDMKCTDCSCKCYEVEWLNQOEAFKOK 400
354 RGDKSKSEYKCSRNKIDCQOTISRKGRKMGKCTDCPFRACSGYEWIMNORQFQKOK 413
401 EYKEKELIASLNDNKEVNNIN- -SEYKQFYEKLETOYATNDTFLNLNEGYCK- - - 455
414 -KTKELISDGGRRKRAVGTTTKEGEKSEFYELKNDGDTGADAFGLLNEKACDIT 472
456 - - - - -GGLPGBKDTITFNSADKGIIFRSEYQVCPQCKGKIDIX 497
473 DGGKINEKYNSSGGVVGSGSGTSGASGNTD- -ENKGTFRSEYQVCPQCKGKIDIX 525
498 THKSDNREHVNEDYKPPAG- - -VKPTN- - -ITVYLSGNQGD- - -TTOKLENC- - -NS 546
526 -HGGNOMER- -KTKYKMKRMSKILKPLNGKMLVLLSKLAKYVKKMKEKCELTQNS 583
547 S- - - - -TNYKDKNNOR- - -WECYKDEINIRKLEONTETIEND- - - - - 581
584 SDGSVGSVTTGASGSEKELYDEKCK- -YKHNVEQKAVVQGEVEDDELKAGAGLCI 642
582 - -NPK- - - - -IISFHNFEELVYTYLLRDTIKNDK- -LTKCINN- -TTTH 620
643 LPPNKKRKEVSEAKSONNHADIQTFHDFYVVAHMLKKSIMHRTKRLKSCISDGKTMK 702
621 CIDEKNCNLCFDRWVYKQKEEWNISIKLFTKKKNIQOASYNNINLFEYGFVMDKLD 680
703 CRGCKAKCCCFEYKAVQKETEEMKPIKDHKTQDGBIEGY- - - - - 743
681 KDEAKKMELENIRKKNNESENNDYLENALELLDLKETATICKDNNTNEACETS 740
744 - - - - -FTTLE- - - - -LILKLOFLKE- - - - -DTEBTEENS 767
741 HNAITNCPVPRGTOPTKAIKEIAYFKRSAYEABNRGILKLGKAHEGITYRGGRK 800
768 LDA- - - - -EEA- - - - - 773
801 DFKONLCRIMIKHSNRLGFSNGPCDGGTQTRFVYGTETWEVDPBHRKDHEDYIM 860
774 - - - - - 773
861 PPRRRHCTSNLEHLOTDDHPLNGNIVDVLVNNSLFDVLLSAKYEANKIIRYKERNL 920
774 - - - - -BELKHLQ- - - - -KILKLENE- - -NLL 791

921 KGPKEVTDPKHOTTICAIIRYSPADIDITIRGRDLMERNQDMVKLOGLHETFGNIHKS 980
792 AVYNAGTE- - -OKTL- - - - -MDKLLNH- - - - - 810
981 KKGNDKYNDAAPRYKLRLRNMMWEANRAKYWEAMKODIKTL- - -KDKSGHSTOSYSGYS 1038
811 - - - - -ELN- - - - -DARKCDPLPEEDSKSRGS- - - - -A 834
1039 DHTPLDYOIPOKLWMTWEMAEWYCKYQKKEVDKLEKCKECKDKDKNOGCTKESGTCTK 1098
835 DHP- -DIF- - - - -RPEKEDDE- - - - - 852
1099 CTEACNEYNDIIGLWKEQWNIISDKYKELHQAQMSYNSGSEASTAKNHNIDRNVIETL 1158
853 - - - - -NEDDD- - - - -EDEVHDEBETAKETTESATDTTSL- - - - - 883
1159 SELYQNGGKSNKSGTSDSAVIGTNTYENVAGYLAHDTGNFDDCQGNFCDEKSDGKD 1218
884 - - - - -DVCPYGVKLT- - - - - -KD 896
1219 NEKYAFRDKPODHGACGCKSGSKPTFVQIKTKKAKBEKDECTVNDILKENDGKKQVE 1278
897 NESL- - - - -QD- - - - -ACSLKYGNNSLRGWRCVTPSGEPTT- - - - - 928
1279 DCHPKKSNNGYPMQCGNINLVEDPRVCMPPRROKLCVHFL- - - - -ANDNEIKK 1327
929 - - - - -SDKNG- - - - -AICVPPRRRLYIKYIWMATKTESPOASGSEASS 969
1328 L- - - - -OSOVNLKEAFIKSAAEFPFSWYKY- - - - -SKDEGNEL- - - - - 1363
970 TSGSTTPPDSKELALKKAFVESALETFLMRKYEKKAAQABAGLPRVEGSEYD 1029
1364 -DKELKGGKIPPAFLRSMFYTFGDRDLFTGDISKGGHSGSKLEQIDSLFKNGDKSP 1422
1030 PEDKLEKGIIPDGLRQMFLLTGDRDILF- - - - - 1059
1423 NGKTRQEMWTEHSEIETEMALCALVYIGAKKDEFTENYGINNAKFSKSTLLEFAKRPQ 1482
1060 - - - - - 1059
1483 FLRWLTWEDDYCYTRQKYLKDVQEKCKSNDQKCTEKNKCKEDYKMKKKKEMIPQD 1542
1060 - - - - - 1059
1543 KYKDERDKRRFRQHIGVWVDTYTGNADYLNRRKFTASGCDKPGSASVYORNIQLEK 1602
1060 - - - - -SGSNDT- - - - -TSVSKDPPSSN- - - - - 1077
1603 QAYTDADKHGCGTYFIENDKYTNISSKDKCKGLYKCANLGAIRKQKGPNTNINKELT 1662
1078 - - - - - 1084
1663 EDVLFPSRLRLICPHALDGNVTDPEVDENGLRKRLMEVATEGYNLGOYKKEKKEKI 1722
1085 - - - - -LASGSTE- - - - -QERKKA 1098
1723 KTSDAHRYSEYVPCSAKSYFYDLRDIILGIDNLEDEKQTEENLKKIFNKNGTSVGK 1782
1099 - - - - -NRY- - - - -KEIKFRKSTERS 1115
1783 SDSTTGNGPSTARKFEYNNKECEYNNAMICGYKRGDRDGSNGSARDEDELKKGSPSD 1842
1116 APNLVSHR- - - - -QTWENNGKXYIMHMYCAL- - - - - 1142
1843 DDYPMGRNDEGTAYOFLRMPFAEWGEDEFCRKHKELEKLYGACNDYTCGDMEDRKKCTD 1902
1143 - - - - - 1142
1903 ACTQYKFEISEWKQYKQIKRYGENKDKIYSEHPVAKADEAREYLDKOLKIKENKSG 1962
1143 - - - - -TSKDKI- - - - -AKGYE- - - - - 1153
1963 DCEYCKAKDVSTORLTLTGNSONMNPASLDDEPKVEYGKCNQVPPRPVRRRETPSPRVSL 2022

Dd	1667	GTCGCKKEKKDE-----	-KK--EESE-----	1668
Oy	3055	WMENNRQIMHAMLCGYKIATSKVTLDEGCWCOLPKPEETNOGLRWLIEWAKOACKERKHV		3114
Dd	1687	-----	PAEESG-----PAEEPAPT	1702
Oy	3115	SDSLTKCPRSNEDNEFASELLRPQCGQNDIRKYSILNLIKMTMENLIKVKOLKOQSS		3174
Dd	1703	AASEET-----ETNP------BPG-		1717
Oy	3175	GNIDNKPSBEENVOSYISKSDQCALBLNDINELVTGTKNNNENEKEVIKKLYPGLPYVE		3234
Dd	1718	-----	TGPAPPTPAPT-----	1731
Oy	3235	DETHKNHVLGDNIKEEQTVPRKALYFTTPHVDSFYQAPLFSTHRVAQYDPRKDILKSI		3294
Dd	1732	-----PDTPPLRPOA-----DEFPDST-----ILOPTI		1755
Oy	3295	SVIVYSVALGLIALHFMKRKSSV-DLLRLINIPQEGYGMPTLESKNRYIPYSGPYKGK		3353
Dd	1756	PFGVALMLAGIAFLFKKTKKASVGNLPQLIOPKSDYDIPTLKSSNYIPYVDRAKKG		1815
Oy	3354	TYIMEGDTSGDECKYKWWMDLSSDDISSESXEELDINDIYPGSPKKTILEVYLEPSK		3413
Dd	1816	TYIIMEGDS--DEDKTAFMSOTDYVSSESEYEELDINDIYPGSPKKTILEVYLEPSG		1873
Oy	3414	RDIPSD--DTPS-----NDTPRTNRFFIDDMMNELKDFVSOYLPTNEPN---NNTKSA		3461
Dd	1874	NNTLASCKNPSPDRDINDIQNDIGPSSKITDNEMNQDLKKEFIISMNLQN-QPDVYNQDTSC		1932
Oy	3462	DIPMTE-----PNTLSDNDEEKPFITISHDRDLTGKEISTYNIMSTNTN		3508
Dd	1933	NSSTNTNITTTSHRNVDNNTNTMTSMNDNMENMLPSIHGNTLYSGEYSYVNM-VNSM		1991
Oy	3509	NDIPMANANDSYRGIDLINDSL		3530
Dd	1992	NDPIPEDNNVYSGIDLINDSL		2013
 RESULT 12 AAW00385 standard; Protein; 1726 AA.				
XX	AAW00385;			
XX	21-FEB-1997	(first entry)		
XX	Truncated Plasmodium falciparum erythrocyte membrane protein.			
XX	Plasmodium falciparum; erythrocyte membrane protein; malaria; detection; identification; treatment; prevention; parasite.			
XX	Plasmodium falciparum MC type.			
XX	Key	Location/Qualifiers		
XX	Domain	62..394	/label= Duffy binding ligand domain 1	
XX	Region	607..648	/note= "Cysteine rich motif"	
XX	Domain	839..1282	/label= Duffy binding ligand domain 2	
XX	Region	1488..1523	/note= "Cysteine rich motif"	
XX	MO9633736-A1.			
XX	31-OCT-1996.			
XX	26-APR-1996;	96WO-US05798.		
XX	27-APR-1995;	95US-0430908.		
XX	(AFFY-) AFFYMAX TECHNOLOGIES NV.			

XX Baruch DI, Howard RJ, Pasloske BL;
 PI WPI: 1996-497376/49.
 DR N-PSDB: AAT41853.
 XX
 PT New Plasmodium falciparum erythrocyte membrane proteins - used to
 PT develop products for the diagnosis, treatment or prevention of
 PT malaria parasite infections
 XX
 PS Claim 1; Figure 12; 149pp; English.
 XX
 CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
 CC membrane protein 1 (PfEMP1) or active fragments or analogues of that
 CC protein can be used in the treatment or prevention of symptoms of a
 CC malaria parasite infection. The polypeptides can inhibit, block or
 CC reverse the sequestration of erythrocytes in patients suffering from
 CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
 CC probes and primers to identify a Plasmodium falciparum parasite, the
 CC primers used to generate characteristic amplification patterns from
 CC different P. falciparum strains. Antibodies specifically
 CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
 CC used in diagnosis of malaria infection. This is a truncated PfEMP1
 CC protein of the MC type of Plasmodium falciparum. The full length
 CC PfEMP1 protein is given in AAM00384.
 XX
 SQ Sequence 1726 AA;
 Query Match 6.4%; Score 1244.5; DB 17; Length 1726;
 Best Local Similarity 22.2%; Pred. No. 1.4e-65;
 Matches 501; Conservative 276; Mismatches 632; Indels 843; Gaps 88;

QY 33 KSARNVLEARYAKNI-----RHPSYAKEHVDLSKGLDTKAEFRGSGSPFNKHNYPPY 86
 DB 13 KDAKHALDRIGEYREKYEYENDAEYKK-----ALKGNLQAKGIGELASS-----PN 60
 QY 87 PCMLDKREHNINLRYDDVNLRHPCHGROQNFDEDESECC-NKIRYKRNDAI-ACAP 144
 DB 61 PCFL-VEEYNNRLK--RKRYPCANROYAFSEYGGQCTFNRIKSENNDSLGACAPY 117
 QY 145 RRRHMDKNLEALNDINTONIHDLGNVLTAKEGESIYNH-PIHGT-----SDACTA 198
 DB 118 RLHLCDYLNLEKMGKSTTK-HGLLDVCAAAATYBGDSITHTTKHLLTPDRKSQCTI 176
 QY 199 LARSFADIGDIVRGIDME-----KPNVHDVETGLREVEFKIHGMEDEVKNDYDPGS 252
 DB 177 LARSFADIGDIVRGIDLYGLYDDEKDERKLLNNLEIFKIHENLGTDADHYKDE 236
 QY 253 GNTYKLEAMNNVNRKMYEATTCDSYKSGYMOSESNTPLSNP-KCGHKQG----- 306
 DB 237 ENTYYQLEREDWVIANRSTVMAITCHAGESDPRFKTCGSGEWTDDKCRCKDEGKNETNE 296
 QY 307 VPTNLVDVPOYLRFMDEMGEEFCRKRNIKLKVKYDSCRNK--ERYLCSHNGHDCITIM 364
 DB 297 VPTFYFDVPOYLWMEEMADDFCRKRKKIENAIKNCRGKGNR-YCDLNGINCEETAR 355
 QY 365 KKGILHLDNKCCTDSTCKKFEYWLGNQDEAFKKOKKEKYEKIO-----SYLSDNKEVNN 420
 DB 356 GAELFVKGDDCHRCVACDFVWIMDQKREPKOKKKYDEINKTGTITTTGNGKINN 415
 QY 421 INSEYKQFKEKLEKETOYATNDFLNLNGKCYK--GGIPGEKIDITFTSADDKGIFR 478
 DB 416 L--YVGHFPIIKL-YTPYVDSLOKLNDLDAICKPPNNGNKASVVDNNEENTTFSH 471
 QY 479 SEYGVVPCDGVKCD-----GIKTYHKS-DNDREVNVEDYKPPAGVYPTNITVLYSNGE 533
 DB 472 TTYTCEAPMGAGQEKNGKNGWAKAKESCAKKERIFNKE-----NSTDIKILTPKGR 524
 QY 534 GDITQKLECNSTNYKNNKNOKECYKNDENINRCKLBQNTIIND-----N 582
 DB 525 SKTELEKLTFCCKDQOKIK--NDIMWKCHYDNG-----TDDQTDSDNDVYLGDMGNLTKE 576
 QY 583 PKIISFNFELWVTYLLRDTIKWMDKLTCTIN-NTTTHCIDECNRNCLCFDRWVYKQKE 641

DB 577 DKIMSTNAFWMVWHDMLDLSIKWRDEHGRCLIKDKGKTICIGCNKKJCIFQWVQDKPT 636
 QY 642 ENNSIKKLEFTKKNKIOOSYNSINNLFEGYFFVMDKLDKDEAKKREIMENIRKKRNEFS 701
 DB 637 EMGKIDHFRKOKDIPKDW----- 655
 QY 702 NLENNDYLENAI--ELLIDHLEKATATCKDNNTNEACETSHNATNPVCPKPRGQTPRK 759
 DB 656 ---THDFLQTLTKMLDLEIITDLY-----GDANE----- 683
 QY 760 NIKELIQVFRKSAVEEARNGHLKLGKAEGLYKKGGRKDKPDNLCRIMIKHSNRLG 819
 DB 684 -IKRI-----EALLEQA--GVGGIDFPAALAGLYTKGFVAE--KDTTIDKLLOHEOK-- 729
 QY 820 FSNGPCDGKGTGDIQTRFVVGTEWEDPEHMRKHEDVIMPPRRRHICTSNLEHQTDD 879
 DB 730 -----EADKCLKTHDTCPPQ-----ED 748
 QY 880 HPLNGNIVDDLNNSPFLGVDLLSAKYEANKILMYKEKNLKGPKVEYDPRKHOTTCRAI 939
 DB 749 RSV-----ARSESATV-----PSPPADRK----- 767
 QY 940 RYSPADIGDIIIRGDLERNRGMVYKLGHLFYPGNHKSJGKGNKYNDAPKYLKLR 999
 DB 768 ----- 767
 QY 1000 ENMWEANRAKWEAMKCDIKYKDKSGHSTQSSYCGSDHTPLDDYIPQKLRMTEWAE 1059
 DB 768 ----- 767
 QY 1060 WYCKVOKEKYDKLEKCKECKDKDNGCGCYKESGTGCTKTEACNEYNDIILGKMGOWNI 1119
 DB 768 ----- 767
 QY 1120 ISDKYKELHQAOAMSVNSNGSIEASTAKNHDNVIEFLSELYQOANGKNSGTSDESA 1179
 DB 768 -----ATEVDAN-----ASSDE- 781
 QY 1180 VIGNTTYENVGAYLIHDTGNFDDCOQSONEFCDKSDKDNKRYAFRDKPODHGAGCGKS 1239
 DB 782 -----DDFEEREE--EEDDEBEAEAEVQOEKTD--ESATPAVA 816
 QY 1240 GSKPTROYIKTKKAEKDETE-CKYVNDILKENDGKQVDECHPKKNSNGYPMQCGNIN 1298
 DB 817 PSPPGTQDQVKKPASQEDDVKCSYIDKALK---GKLD-DACTLYKGTAPTSMKC--- 868
 QY 1299 LVEDPR-----VCMPPRROKCYVHFL-----A 1320
 DB 869 IPSDTISVATGSDTGTGSGSICVPPRRKKLYVGLHDWAGETTEAKSOETSGGQKTPSG 928
 QY 1321 NDN-----EIKKIQSOVNLKEAFIKSAAEFFPSWYTKSK----- 1356
 DB 929 NESPSPSKLPQGRPTPTTKTPESSLSLHAFVSPPRLRLPLMKFKQMAQAQACATGL 988
 QY 1357 -----DGEENELDKLEKGIIPPAFLRSMFYTFGDRFLFTGD-----ISKG 1399
 DB 989 QLPGVYVDDSDPPOQLTKRGNIPNDFLRQMYTLAGYRDICIGGDRDIDGDTIVSITEG 1048
 QY 1400 HGEGRKLEKQIDSLFKKGGDKSPNGK-----TROVWVHEHSIEIYEA 1442
 DB 1049 ESTRKIKSLIEGLFKQVYSPSPROTSSRTVPVHPQTSVEKTPQQTWMBANGPHIWMG 1108
 QY 1443 LCAALV-----KI-----GAK-----KDDFTENYGVNNVFKSDKS 1471
 DB 1109 ICALTYSAGAIQPPQKVADADKYLEKLKPNANGIKWMLKEDNTSSAMPTSSSSSGS 1168
 QY 1472 T-----TLEFAKROFLMLFEMWYDCTYTRQYKLVADOEKC----- 1510
 DB 1169 NDPINTPKLTFEVEIIPFFRYLHEWGCNFCERKRLKQLYKECKVGENGYGRGRQKTP 1228
 QY 1511 -----SNDOLK-----CDTE--CNKKCEYVYVMKKK-EMTIPQK-----YY 1545

Db 1229 QCSYGEDECDJLSKSYDYADIECPKCAHCHMYKKMIKKKDEFTQEKAPKODKY 1288
 QY 1546 KDERDKRRDR-----QHIGVAVDTYGTNATDYL-----NKRFTASCG 1584
 Db 1289 VGNNGGNGGNGFCITLSTSDAAGLEKISGCKKNSENGNDKLNFSQNETFPATV 1348
 QY 1585 DRGSAVYQVQRIQLEKQAYADRHCGGTGRIENDDKYTNISSDKCKGLYKANTGA 1644
 Db 1349 CRKCSFEKIDCKENCKCKNGGGGTNETGCTGT-----TSENFKQKG-QTAKERY 1398
 QY 1645 IKWQNGPNYNNIKELTEDV-LFPSRL-----RICEHAL-----DGNATDPYVKDE 1691
 Db 1399 MRVSDNNPNPGEFDLNEACQAGIFKIRDEMGCGVCCGYCIVIPKNGCVTTSGENNDQ 1458
 QY 1692 ----NLRRLMEVATGEGNIGQYKKEKEKITSAN-KSYEVPPC----- 1737
 Db 1459 IITIRIVAHWVQ-NFLDYDN-----KIKHKISHCKNSSEGYTCIKNVEQMI 1505
 QY 1738 SAMKSYFDLRDITIGIDNLEDEKTEENIKKI-----FNKN-----GTSVAKGSD 1784
 Db 1506 STKRTWTNIK-ILN-----EYKQNPDIYVNTITLLODLOSQIDFNKAIKPCGT-LTKFED 1559
 QY 1785 STTGNGSTARKFFMNEKNCYNNAMICYKGR--DDGNSGNSARSDDLKTC--GSV 1839
 Db 1560 SSGINGAESSEKKNHGE-----YDAIDCMNLRLQDKIDCONNHNAGNGENAKCEKHA 1614
 QY 1840 PDDDDYPMKKNDEGATAYOFLNFAWGEDFC-KIKKELEKLVGACNDYTCGDNEDKR 1898
 Db 1615 PDDDDDEALEENPVTQ-----PNICPKPEPEKAEKGG-----CEPAEKKEK 1657
 QY 1899 KCTDACTQYKFKISEWKPOYKQIKKYGKNDKIYSEHPV-AKDAEDAREYLDQKLI 1956
 Db 1658 V-----EKEEKTIVNTVAKPTEKEAAGDPAAGFAADSEENP----- 1693
 QY 1957 CENKSGDCYKCKMDYSTORLTGDSQNNPAS 1988
 Db 1694 -EKAPEPEVETKDKAPVKT-----PAS 1717
 RESULT 13
 AAB18144
 ID AAB18144 standard; Protein, 1700 AA.
 AC AAB18144;
 XX 07-NOV-2000 (first entry)
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:1.
 OS Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 XX antimalarial; malaria; protozoacide; infection; insecticide.
 OS Plasmodium falciparum.
 PN WO200025728-A2.
 XX 11-MAY-2000.
 PD 05-NOV-1999; 99WO-US26796.
 PE 05-NOV-1999; 99WO-US26796.
 XX 05-NOV-1998; 98US-0107131.
 PR (HOFF/) HOFFMAN S.
 PA (CARU/) CARUCCI D.
 PA (GARD/) GARDNER M.
 PA (VENT/) VENTER J C.
 XX Hoffman S, Carucci D, Gardner M, Venter JC;
 PI WPI; 2000-365347/31.
 XX Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the

PT diagnosis of P.falciparum infection -
 XX Disclosure; Page 29-33; 577pp; English.
 PS The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (I) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasite life cycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAB70078 to AAB70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.
 CC
 CC Sequence 1700 AA;
 SQ

Query Match 5.4%; Score 1045.5; DB 21; Length 1700;
 Best Local Similarity 16.0%; Pred. No. 1.3e-53;
 Matches 576; Conservative 198; Mismatches 622; Indels 2195; Gaps 83;
 QY 33 KSARNVLERYAKNI-RHPSKYAKEHVDLSKDLTFAEGRGSPFPVKNHNTYV----- 85
 Db 13 ESKVDFRIGKRYEKEKTEKAKRTTELHGLDSATV-----PNDKH-----EKGSTEN 62
 QY 86 YPCNDHEHNRILKRDVYLNKPCGKGRONRDEDESECGNKIRYKRNDAIACAPR 145
 Db 63 NPKLQYDNTNVTN-GEQGEPECTDIYERFSPDEGAQCDK--KIKDNSEGACAPYR 118
 QY 146 RHHMCKMLALNDIN-QONIHDLGNVLYAKYGESEIVNHP-HKGT-----SDCTA 198
 Db 119 RLHVCVRNLENIINDYSKINNNHNLVEYCLAKEGISTRTGYPOHOBTPDYSQLCIV 178
 QY 199 LARSFADIGDIYRGIDMK-----PNVHDEVETGLREVEKIHDM-----EDEYKN 245
 Db 179 LARSFADIGDIIRGNDLYRGNTKEKKKKLEKLTIFGHIIDELKNGTNEEELQK 238
 QY 246 DYNPDGSGNYKRLAEAMVNNVKNVWEAITCDA-SYSGTFMOSESNTPLFSNPKCGHKQ 304
 Db 239 RYRGDKNDYOLKREDWMDANRETVWKAITOMASGYOQPTCGRGELPYVTLKSCOCIA 298
 QY 305 GKVPNLIDYVQYLRWFEDWGEFCRKRNIKLLKVKKSCR--NDKRLYCSHNGHCTT 361
 Db 299 GEVPTIYDYYQYLRWFEMAEDEFCRKKKKIPVKNCRQVOVGKKE-YCDRGVYCDG 357
 QY 362 TIKKGLIHLNDKCTDSTCKVEFWLGMQEAFFKQKEKEKEIOSYLSNDKFFVNNI 421
 Db 358 TIRKQYIYRLDPTCTKSLACKTFAEWIDNOKQEOFKOQYQNEISGGGGRQRKSTHS 417
 QY 422 NSE---YKQFYELKLEQYATND-----TFNLNLNGKCTCKGLP-GEKDITFTNSADD 472
 Db 418 TKKEYGEKHEHNEELR-----NEKGRVSRFLQLSKEKICEKRIQVGEETANGYENNE 471
 QY 473 KGIFYREYQVCPDCGVKCDGIKYTHKSNDRERVANNEDYKPPWGVKPINIYVLSGNE 522
 Db 472 SNFESHTEYCDRCPGLGVDC-----SSDN----- 495
 QY 533 QGDITQKLENFCONSINYKDKNNKQWECYKDENINCKLEQNTETNNDPKTIISPHNF 592
 Db 496 -----CPFB----- 499
 QY 593 ELWVTYLLRPTIKWNDLKTICINNTT-----THCIDBCNANCICFDRWVYQKEEENNSIK 647

Db 500 -WAT-----GGCGACGTGTAGTGGCGCATAGTTTAC----- 531
Qy 648 KLETKKNNIQOYSYNNINLPEGTFEKKVMDKLDKDAKKELMENTIKRKNESNLNNR 707
Db 532 ----- 531
Qy 708 DYLENAIELLDHLKETATICKDNNTNEACETSHNATTPNCVPRGTOPTKNIKEIAQY 767
Db 532 -----GGGATCAAGGTGTCAAGATTAT----- 555
Qy 768 FKRSAAEARNRGLHLKGAHGIYKRGRRDFKDLNLCRIMIKSHNRMLGFSNCPDG 827
Db 556 -----TTGATAG 562
Qy 828 KGTGGGIGTREFVGTMEVDEPEHRRKHEDVIMPRRRHICTSNLEHLQTDHPLNGIV 887
Db 563 AATAGGGAAGAAAGT----- 577
Qy 888 DDLYNNSLFDVLLSAKYEANKIIRMYKEKNNLKGPKEVTDPKHQTTICHAIRYSPADIG 947
Db 578 -----TTACGAAAAAAC----- 589
Qy 948 DIIRGHDLMERNGDVNLQGLHLETFEGNIHSLKGNDRKNDAPRYLLKRENNWEANR 1007
Db 590 -----AGA 592
Qy 1008 AKYWEAMKCDIKYLKDKSGHSTOSSYCGSDHTPLDDYIPOKLRMTWMAEWYCKYQK 1067
Db 593 AAAGATTGC-----AAAACGAT----- 609
Qy 1068 EYDKLEKKECKD-----KONGOGCTRESGT-----GCTGC-TEACNEYNDIGIMKE 1115
Db 610 -----ATACTACTGAATGTGATGTGAT-TTGCAAAAAGAACATATC----- 651
Qy 1116 OMNIISDKYELHEOAMVSNSGIEASSTAKNHDNRNVEFLSELYOQNGSKSNKSGTS 1175
Db 652 -----CAATGATAAACAT-----CCGAAGAGATCA 678
Qy 1176 DESAVIGTNTTYENVAGAYLHDGTGNFDCCOSQNEFCDEKSGKDNKAYAFDKPOHDGAC 1235
Db 679 CAGAAATATATCCATG-----CAAACTCATATATGATTATATCTAATGTTAC 727
Qy 1236 GCKSGKPRVQJIKTKKAEEKDECKTVNDILKENDGKQOVEDCHKKNKSNVCPMDQCG 1295
Db 728 TCAATGTTT----- 737
Qy 1296 NINLVEDPRVCMPPROKLCVHFLANDNEIKLQSVNLKEAFIKSAAETPFSWYYS 1355
Db 738 ----- 737
Qy 1356 KDEGNELEKELKGIIPAPFLASMTTFGDIYDFLEGTDISKGHSGSLKQIDSLFK 1415
Db 738 -----GGTCAAGAGAT----- 749
Qy 1416 NGQKSPNGKTRQEMWTEHSHEIWEAMLCALVIGAKKDFTEHYGVNNKFPDQKSTLE 1475
Db 750 ----- 749
Qy 1476 EFAKRPQFLMWLEWYDYCYTRQKYLKQVQCKSNQDLCDTECNKCEDYVYKMKK 1535
Db 750 -----CCT----- 752
Qy 1536 KEMIPQDKYKDERDKRRPRQIHGVWVTYTGNNATDYLNKRFYASCGKPPSASAVOR 1595
Db 753 -----TGGA-----AACGACATAGT----- 769
Qy 1596 NIQLEKQAYYADRKHCGCTKFIENDDKYTNNISKQCKGLVKEANNGALKNQKGNNY 1655
Db 770 -----AGAAGCTTT-----TTGATACAGAGAG----- 795
Qy 1656 NMLKELEDEVLPSSRLRILCFHALDGNVTPDEVKENGRLKRLMEVATGEGYNIGQYKE 1715

Db 796 -----CACATGATATA-----GAAAAAATAAAGATA----- 826
Qy 1716 KKEKELIKTDAHKYSEYVPPCSAMKISFYDLRDIIGIDNLEDEKOTEENLKKIFNNK 1775
Db 827 -----TA 828
Qy 1776 GTSVKGSDSTGNGPOSTARKFFMNNENKECVWMAIMCGYKRGNDGNSGARSDEDLK 1835
Db 829 GTGAGAGAGCTTG-----CG-----CTC 846
Qy 1836 CGSVPSDDDYPMKRNDEGTAYQFLRWFAMGDEFCFKHEKELEKLYGACNDYTCGDNED 1895
Db 847 CATATA-----GAG----- 856
Qy 1896 KRRKCTDCTQYKFIJSEMPQYEQIKKYENKDKIYSEHPVAKDAEDAREYLDQK 1955
Db 857 -----ATTACA----- 862
Qy 1956 ICENKSGDCEYKCMKQVSTORLTDGNSQNNPASLDBDEPKVEGKNCQVPRGPRVRRRT 2015
Db 863 -----TGTAT-GCGTTAGAAATTGG----- 882
Qy 2016 PSPRVSLSKATASKKEAKTAPPTKQPKYENLTTEMRQOTRTRAAOQTRKRTSTATTT 2075
Db 883 -----AAATATCATATATTA-----TACTAAATTAATAATAAACATATAT 924
Qy 2076 ESDVGTWKAILSNKPRGIEGECNPKTYGQYPRKMGCIYKSKENENGICMPRRRKLK 2135
Db 925 TAVTGTAGAGAGGTGTCTTGACAG-----CAATATGAGAG----- 961
Qy 2136 INNIOYLVTEENKRDNDIEAFIKCAIETQFLMLKYLIIENPAEENLQNTIPDEFKR 2195
Db 962 -----GGAATCAATA-----CAGTC----- 978
Qy 2196 IMYTYGDKMEFGTDISNDKLIYVNSVTLLNENNRKODKKDEELRKIFWEK 2255
Db 979 -----GT-----TKTDROQF-----FNGTYG 995
Qy 2256 KFIWEGHLYLTHLIDENKEKIRNYQYNDMT---KLT-PSLEEYVKKRQFLRMFTW 2311
Db 996 KDIWKMLCA-----LOEAGGKTLTETYNYSNVTFNGLGTKLNEFASPSFLRMFTW 1051
Qy 2312 AEEFCNRKROLLKLEAGCEYBENG--SNDGKIOEACAVYQNFIRKMTKEFEQRE 2369
Db 1052 GDQCFREIRITQOLILKRCWYQYNGDKKODKKEKTECTYKEMLTWMDYKQNO 1111
Qy 2370 KFKKDKKKYKDYPSSTERDIEKATCAHEYLNNMLKE-LCGND---CSCMQKPSQLP 2424
Db 1112 RYEVKGTSPYKE---DSQVKEKVAHGLYLRKILKNIICTSGDIYACNME----- 1160
Qy 2425 KTTQOSQSDANDMPESLDYVPEEFNK-CEQPELSKSGSMHTKTKTEKIPMNCVEKAA 2483
Db 1161 -GSTSDSSNNNDNIPESLKPPIEIEGCTCKQPS-GEVYIPEKKYVEPRV----- 1209
Qy 2484 YVLSKEAENMDITLKKFPIPIESTKEKSNSTNNPPDKPKPAPADYIGRRNCEN 2543
Db 1210 ----- 1209
Qy 2544 REENRFVDEYEMKYCKNSKEYOEKRRVCVPRRREHMLRLNDEIKIERLDSNYLLKMR 2603
Db 1210 ----- 1209
Qy 2604 RTARNESIDILKFNSENGCANP ICDTMYKYSFADLGIYRGDMLRIGIYLPVEIKLY 2663
Db 1210 ----- 1209
Qy 2664 KVEFYIYKWRNKKGNKYNDVQTFPSAMWANDRKDIWKAMCKAPEDAKLPRKGMGD 2723
Db 1210 ----- 1209
Qy 2724 FERITLLQDKGHRDPPVDYIIPQRRMTWSEYCYKALMELEKFKKSDCHCKTSR 2783
Db 1210 ----- 1209

```

OY 2784 CNDYDENKCEOCKTCQGEYKMFVLMKSLFDIOSNKYKELYEOPITYRTISTYDHVONFV 2843
DB 1210 ----- 1209
OY 2844 OKLTKFKSCSVSESEYELHETSKCLNTYFNENDSSNIRTYAFETTPKSYKACSCITP 2903
DB 1210 ----- 1211
OY 2904 SKNPNDNCPDONKCKELQTFETCSKNDYNNLDNMNAYLVNNSDDNKGVILPPRRR 2963
DB 1212 K-----PPK----- 1215
OY 2964 HLCRTRITAYNRKGDKEILKLLTSAPSGQLGQTKSEEECEFAKMSYADYSOI 3023
DB 1216 ----- 1215
OY 3024 IKGTDMDTLSSEKIKIIPETSEATENRKTWMENNRQIMWAMLCGYKIATSKYTLDEG 3083
DB 1216 ----- 1215
OY 3084 WCQLPDEBETNOFLMVLIMAKOCKEKKHVSISLTKCPSRNEEDNFEASELLRQPGCQN 3143
DB 1216 ---LTK-----ROP----- 1221
OY 3144 DIRKTIISLILIKNTMENLNIYKOLKDOSSGNIDNKPSSENVOSTYKSKDQCALEND 3203
DB 1222 -----KERD----- 1225
OY 3204 INEIVTGKNNENNEFEVYLKLPGLYVEDETHKNHVDGNINKEBQTVAPKALYFT 3263
DB 1226 -----PPT----- 1228
OY 3264 PHVDSFYQAPLEFSTRHVAQYDKNDILKSSISVIVSALGLTALHPMKKFKSVDLARI 3323
DB 1229 PAL-----KNMLSTIMSIGIFATPTTYFKLKKTKSTIDLLRY 1269
OY 3324 INIPGEXMPTLESKNRYIPRSGPYGKTYIMEGDT-----SGEDKYMIDSSDIT- 3379
DB 1270 INIPSSDDIPKTLSPNRIIPTSGYKCKRYILEGSGDSTGYTDHY-----SDTTS 1323
OY 3380 SSESSEYELINDIIVVPSPKYKTLIEVLEPSKRDIPS-----DDTPS-----NDTPRTN 3430
DB 1324 SSESSEYELINDIIVAPAPKTKTILEVLEPSGNNITASGNNTSPDONIDONGIERS 1383
OY 3431 RPIDDEWELKHDEVSQYLPNTENP-----NNKYSADIPMTEPNTLYSDNPEEKPTIISIH 3487
DB 1384 KITDENWNTLKDEFISQYLOSQEPNDVPNDYSSGDIPLNTOPTLYFDPNDEKPTIISIH 1443
3488 DRDIYTGKEISYNINMSTNTNNDIPMANRDSYRGIDILINDSLVYLVNLIY 3538
DB 1444 DRDLISGEYSINVM-VNTNNDIPISGKNGTSGIDILINDSLNSNNVDIY 1493

```

```

PF 07-SEP-1994; 94WO-US10230.
XX
PR 10-SEP-1993; 93US-0119677.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
XX
DR WPI: 1995-123427/16.
XX N-PSDB; AA083528.
XX
PT New erythrocyte binding domain polypeptide(s) - isolated from
PT Plasmodium binding proteins, used in diagnosis, treatment and
PT prevention of malaria
XX
PS Disclosure: Page 56-57; 81pp; English.
XX
CC Erythrocyte binding ligand (EBL) family genes were cloned from
CC P. falciparum chromosome 7 subsegment libraries constructed during
CC genetic studies of the chloroquine resistance locus. The 4 genes,
CC EBL-el (AA083526), E3la (AA083527), EBL-e2 (AA083528) and Proj3
CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
CC binding domains of such proteins can be expressed e.g. in E. coli,
CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
CC cells, and provide protection against P. falciparum.
XX
SQ Sequence 700 AA;

```

Query Match 5.38; Score 1032.5; DB 16; Length 700;
Best Local Similarity 34.88; Pred. No. 2e-93;
Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

```

OY 136 NDALCAPRRRHODKRLALNDINONHDLGNVLYTKGESSTVNNH--PHGT- 192
DB 5 NKVGACAYRRLHLCIDYLESID--TSTTKLLEVCMAAKRENSI-NHYHQHQTN 61
OY 193 ----SDACTALARSFADIGDIVRGIDMF-----KPNVHDKVETGLREVFRIHGMDE 242
DB 62 BDSASQLOCTVLARSFADIGDIVRGIDLYLGYDNKEQKRLKQKIDKIFKIH---KDV 118
OY 243 VKND-----VMPDG--SGYVYKLRKEMVMVNNKWEATITCASTKSGTFMOSNTPLFS 296
DB 119 MKNMGAERYTDDAKGDFQLREDWMTSNREYWKALICHAPFANFTTACNVGKT 178
OY 297 NPKGSHKQKVPYPLNDYVPOYLRFWDEWGEFCRRNRKIKLKVDSKNDKERLYCSHG 356
DB 179 NGGCHICIGDVPYTFDYVPOYLRFWEBAEDPCRRKKKLENLQCKRDYQNLTCGSG 238
OY 357 HDCTTTIKKGIHLHLDNKCITDCSTCKVYEWALNQDAFRRKREYKEI----- 407
DB 239 YDCTRTIYKKGKLVIGEHCTNCVWCRMYETWIDNOKREFLKQRRKETEISGGSGSP 298
OY 408 -----QSLSDNKFVNNINSEYIKQFEKLEKFOYALNDFLNLNDEGKYCKG--- 456
DB 299 KRTIRARSSSSSDN-----GYESKFFYKLLKEVGYODVDFLILNKEGICOROPQ 350
OY 457 -GLPGEKIDFTFNSADDKGIFRSEYCOVCPGCVKCGIKYTRKSDNDRVRVNNDEYKP 515
DB 351 VGNKEKADNVDFTNKRYK--TFSRTEICEPCWCLGEGGPPMKYKDKTGSAATKYTDP 409
OY 516 PWGVKPTNIVLYSGNREGDITQKLENFCONSINYKDKKNOKMECYKYD-----ENINRC 570
DB 410 K--NIDIDIPVLPDKSOONILKRYKNCEKGAP--GGQIKWKQCYDEHRPSSKNNNC 465
OY 571 KLENTLEINNDNPKIISFHNFEELWYTYLLBDTIKAMDKLCTCINNNTT-----HCIDECN 626
DB 466 VEGTWDFKFTQKQYKSYNVFMDVMDKLDHSDVEKTELSKCIINNNTNGTCRRNNRKC 525
OY 627 RNCLCFDRWYKQKEEWNISIKKLETKRKNIOQSYYSININLFEQYFPKVDKLDKDEAKV 686
DB 526 TDCGCFQKWEYKQOEWMAIKDHGKQTDIYQO-----KGLIVSPYGVLDIV-----LKG 576

```

QY 687 KEMENIKRKNKNEFNSLENNRDYLE-----NAIELLDHLKETATICKDNN 732
 Db 577 GNLQNIKIDVHGDYDDDKIKHKLDEEDAVAVYLGKNDNTTIDKLLQHEKQOAEQCKOK- 635
 QY 733 TNEACE-----TSHNATTNPGVPRGTOPTKNIKEIAOYFRSAAYEE 775
 Db 636 -QEBCCKKAQOESRGRSAETREDERTQ---QPADSAGEVEEEDDDDDYDE 681

RESULT 15

AAW22481
 ID AAW22481 standard; protein; 700 AA.

XX AAW22481;

XX 07-OCT-1997 (first entry)

XX Plasmodium ebl-2.

XX DBL gene family; SABP; stialic acid binding protein; vaccine; therapy;
 XX Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 XX DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 XX Plasmodium.

XX Plasmodium falciparum.

XX WO9640766-A2.

XX 19-DEC-1996.

XX 07-JUN-1996; 96MO-US09508.

XX 07-JUN-1995; 950S-0487826.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Chittis C, Miller LH, Peterson DS, Sim KL, Su X;

XX Wellens TE;

XX WPI; 1997-052231/05.

XX N-PDB; AAT72896.

XX New malaria vaccines - contains cysteine-rich DBL family protein

XX binding domains homologous domains of the Duffy and stialic acid

XX binding proteins

XX Disclosure; Page 46-48; 96pp; English.

XX This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to
 XX the Duffy binding like (DBL) family of genes which have homology to the
 XX Duffy antigen binding protein (DABP) and stialic acid binding protein
 XX (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 XX var family of genes modulate cytoadherence and antigenic variation of
 XX Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 XX protein (DABP) are soluble proteins that appear in the culture
 XX supernatant after infected erythrocytes release merozoites. DABP and
 XX SABP mediate the binding of merozoites and schizonts to the erythrocyte
 XX surface. These proteins are necessary for erythrocyte invasion by the
 XX parasite. This sequence can be used in the compositions of the
 XX invention. The compositions are for the treatment and prevention of
 XX malaria, and comprise either a nucleotide sequence or encoded polypeptide
 XX of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 XX family of genes having homology with conserved regions of DABP and SABP.
 XX The compositions are used for the treatment and prevention of malaria.
 XX They are also used in the preparation of vaccines for inducing a
 XX protective immune response in a mammal to Plasmodium merozoites
 XX (especially Plasmodium falciparum or Plasmodium vivax).

XX Sequence 700 AA;

Query Match 5.3%; Score 1032.5; DB 18; Length 700;

Best Local Similarity 34.8%; Pred. No. 2e-53;
 Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDALACAPRRRRHCKDNLKLEALNDINTQNIHLLGNLVYAKYGESEIVNNH--PHKGT- 192
 Db 5 NKVACAPRYRLHLCIDYNLESID--TSTTHKLLLEVCMAKYGENSI-NHYHQHRTN 61
 QY 193 -----SDACTALARFALIGDVRGIDMF-----KPNVHDKVEFGLAEVFRKKIDHGMDE 242
 Db 62 EDSASQLCTVILARFSAIGDILVRKGLDLYLGYNKEKBOBKLEOKLDIFKKH---KDV 118
 QY 243 VKND-----YNPDG-SGNYYKLRAMNNVNNRKNRYEAITCDASKSGFMOSESTPLEFS 296
 Db 119 MKTNGAERIYIDDAKGDGDFQLRREDMWT SNREYWKALICHAPKEANYFITACNKGKT 178
 QY 297 NPKCGHROGAVPTNLDYVPOYLRFWDEMGEEFCRRRNKIKLKVYDSGRNDRERLYCSHG 356
 Db 179 NGQCHCIGDVPYTFYDVPQYLRFWEEMAEDEFCRKKKKLENILOKCRDYEQNLVCSGNG 238
 QY 357 HDCTTWMKGIHLHDKKCDSCSKCYFEVYVLGNOQAEAFKKOKEKEKEI----- 407
 Db 239 YDCYTTLYKRGKLVIGHCITNCISWCRMYETWIDNOKREPLKORRYETEISGGSGSKSP 298
 QY 408 -----QSYLSNDNKFVNINSEYKQFEKLETOYATNDPFLMLNKGKCKG--- 456
 Db 299 KRTKRAARSSSSSDN-----GYESKFYKLLKVEGYQDVDFKILINKKEGICQKOPQ 350
 QY 457 -GLPEKDIPTFTNSADKGIYRSEYQVCPDGCYKCDGKIKYTHKSDNDRERNEDYKP 515
 Db 351 VGNKADNVDFTNKRYK-FESRTEICEPCPWCGLKRGPPMKYKGDTCGSAKTKYDP 409
 QY 516 PMGVKPNINIVLYSGNQGDTQKLEFNCNSSTYKKNKNNKWCYKQD-----ENINRC 570
 Db 410 K---NITDIPVLPDKSQOQLTKRYKNFCEKAP-GGGQIKKWQCYDDEHRPSSKNNNC 465
 QY 571 KLEQTEINNDNPKIISFHNFEFLMYTYLBDTIKMDKLTCTCINNTT---HCIDCN 626
 Db 466 VEGTWDFKFTQCKQVYKSNVNFEMVADMLDSVEMTELKSCJNNNTNGTCRNNKCK 525
 QY 627 RNCICFDRAWYKQKEENNSIKLFTKRRKNIQOASYYSNINLFEQYFPKVMKLDKDEAKW 686
 Db 526 TDCGCFQKWEKKQOEWMAIKDHFGKQTDIVQ---KGLVFSPYGVLDLV---LKG 576
 QY 687 KEMENIKRKNKNEFNSLENNRDYLE-----NAIELLDHLKETATICKDNN 732
 Db 577 GNLQNIKIDVHGDYDDDKIKHKLDEEDAVAVYLGKNDNTTIDKLLQHEKQOAEQCKOK- 635
 QY 733 TNEACE-----TSHNATTNPGVPRGTOPTKNIKEIAOYFRSAAYEE 775
 Db 636 -QEBCCKKAQOESRGRSAETREDERTQ---QPADSAGEVEEEDDDDDYDE 681

Search completed: April 28, 2003, 10:29:57
 Job time: 224.055 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 132.729 Seconds

(without alignments)
5498.552 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407
Sequence: 1 MGESCKYFIKMGNAASSLE.....IDLINDSLVNLILYIMMKY 3542

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

1 number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	19407	100.0	3542	5	Q9U5M2 plasmodium
2	12954.5	66.8	3287	5	Q8T326 plasmodium
3	7367	38.0	1615	5	Q8T325 plasmodium
4	4515.5	23.3	2706	5	Q15870 plasmodium
5	4254	21.9	3006	5	Q26032 plasmodium
6	4010	20.7	3026	5	Q26030 plasmodium
7	3489	18.0	3078	5	Q26031 plasmodium
8	3328.5	17.2	2527	5	Q95W83 plasmodium
9	3130	16.1	2924	5	Q25733 plasmodium
10	2932	15.1	1685	5	Q9U4A2 plasmodium
11	2869	14.8	2658	5	Q8T360 plasmodium
12	2817	14.5	510	5	Q8T610 plasmodium
13	2714.5	14.0	2664	5	Q26033 plasmodium
14	2686.5	13.8	2647	5	P90580 plasmodium
15	2677	13.8	2212	5	Q94657 plasmodium
16	2556.5	13.2	494	5	Q8T6K7 plasmodium

17	2480.5	12.8	492	5	Q8T6K9 plasmodium
18	2449.5	12.6	494	5	Q8T6K0 plasmodium
19	2441.5	12.6	465	5	Q8T6K4 plasmodium
20	2436	12.6	2710	5	Q9XZB8 plasmodium
21	2413.5	12.4	494	5	Q8T6K5 plasmodium
22	2412.5	12.4	492	5	Q8T6K6 plasmodium
23	2403.5	12.4	496	5	Q8T6K8 plasmodium
24	2335.5	12.0	461	5	Q8T6K3 plasmodium
25	2303	11.9	2135	5	Q61077 plasmodium
26	2258.5	11.6	460	5	Q8T6K1 plasmodium
27	2056.5	10.6	1327	5	Q9NFB4 plasmodium
28	2055.5	10.6	427	5	Q8T6K2 plasmodium
29	1878.5	9.7	2169	5	Q97312 plasmodium
30	1750	9.0	2163	5	Q9NFB6 plasmodium
31	1749	9.0	2197	5	Q96296 plasmodium
32	1743	9.0	2182	5	Q26034 plasmodium
33	1716.5	8.8	2209	5	Q97324 plasmodium
34	1697.5	8.7	2227	5	Q9U0G5 plasmodium
35	1689.5	8.7	2228	5	Q60991 plasmodium
36	1611	8.3	2209	5	Q9U0G6 plasmodium
37	1574	8.1	1711	5	Q96108 plasmodium
38	1523.5	7.9	2042	5	Q25766 plasmodium
39	1249	6.4	1729	5	Q25734 plasmodium
40	924	4.8	173	5	Q96735 plasmodium
41	811.5	4.2	431	5	Q96294 plasmodium
42	749.5	3.9	2647	5	Q9U4X0 plasmodium
43	660.5	3.4	438	5	Q9Y1N7 plasmodium
44	651	3.4	118	5	Q96450 plasmodium
45	646.5	3.3	455	5	Q9Y1N6 plasmodium

ALIGNMENTS

RESULT 1

Q9U5M2 ID Q9U5M2 PRELIMINARY; PRT; 3542 AA.

AC 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DI 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE FCR3 CSA ligand (Fragment).

GN VAR.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]
SEQUENCE FROM N.A.

RC STRAIN=FCR3;

RX MEDLINE=2006305; PubMed=10535993;

RA Buffet P., Gamain B., Scheldig C., Baruch B., Oishi S., Fujii N.,

RT sulfate A: A receptor for human placental infection."

RT sulfatase A: A receptor for human placental infection."

RT sulfatase A: A receptor for human placental infection."

RL Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).

DR EMBL; AJ133811; CAB59840.1; -

DR InterPro: IPR001219; Neurotoxin.

DR InterPro: IPR004258; PFEEMP.

DR Pfam: PF03011; PFEEMP. 1.

DR PRINTS; PR00284; TOXIN.

FT NON_TER 3542

FT SEQUENCE 3542 AA; 413089 MW; 970D85EE8BDA2BC2 CRC64;

Query Match 100.0%; Score 19407; DB 5; Length 3542;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3542; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGESCKYFIKMGNAASSLEGDASKSPIIKESHKSARVLEERRAKNIRHPSYAKAEHVDL 60

DB 1 MGESCKYFIKMGNAASSLEGDASKSPIIKESHKSARVLEERRAKNIRHPSYAKAEHVDL 60

QY 61 KGDITFAERFGSPFPVKNHNTYPPPCULDKREHNLKTYDDVNLHHPCHGRQNNFDD 120

Db 61 KGLTLAEFRGSGSTPVNKNHNYYPVPCNLDHKEHTNLRYDYNLRHRCGRQNEPDED 120
Qy 121 ESEECGNKIRNYKRNDAICAPRRRHMCDDKMLEALNDINTONIDILGNVLTAKYES 180
Db 121 ESEECGNKIRNYKRNDAICAPRRRHMCDDKMLEALNDINTONIDILGNVLTAKYES 180
Qy 181 ESTVNNHPHKGTSIDACTALARSFADIGDLYRGIDMFKPNVNDKVEITGLREVEFKIIDGME 240
Db 181 ESTVNNHPHKGTSIDACTALARSFADIGDLYRGIDMFKPNVNDKVEITGLREVEFKIIDGME 240
Qy 241 DEKNDYNDPGSGNYYKRLREAMNNVNRNYWEALITCDASYSKYPMQSSNPTPLFSNPKC 300
Db 241 DEKNDYNDPGSGNYYKRLREAMNNVNRNYWEALITCDASYSKYPMQSSNPTPLFSNPKC 300
Qy 301 GHRQGVPTNLDVYPOYLRFWEDMEGEFCRKRNIKLKYKDCSRNKEKELYSCHSHDCT 360
Db 301 GHRQGVPTNLDVYPOYLRFWEDMEGEFCRKRNIKLKYKDCSRNKEKELYSCHSHDCT 360
Qy 361 TTTMKGILLHDKKCDPSCSTKCVFVWILGNOOBAFKKOKREKYEKIOYLSNDNKFVNN 420
Db 361 TTTMKGILLHDKKCDPSCSTKCVFVWILGNOOBAFKKOKREKYEKIOYLSNDNKFVNN 420
Qy 421 INSEYKOFYEKLKETOVATNDTFLNLNEKRYCKGGLGEGEDITPTNSADKGIIFYSE 480
Db 421 INSEYKOFYEKLKETOVATNDTFLNLNEKRYCKGGLGEGEDITPTNSADKGIIFYSE 480
Qy 481 YCOVCDGCVKCDGKYTHKSDNDRERNVNEYKPPWGVKPTNTIVLYSGNEQDITOKL 540
Db 481 YCOVCDGCVKCDGKYTHKSDNDRERNVNEYKPPWGVKPTNTIVLYSGNEQDITOKL 540
Qy 541 ENFCNSSTNYKDNNOKWECCYKEDENINCKLEONTLEINNDPKIISPHNPELWVYLL 600
Db 541 ENFCNSSTNYKDNNOKWECCYKEDENINCKLEONTLEINNDPKIISPHNPELWVYLL 600
Qy 601 RPTIMNKLKTCINNTTTHCIDEENRNCLEPDRWVKOKEEWNSIKKLTFTKKNNIOQSY 660
Db 601 RPTIMNKLKTCINNTTTHCIDEENRNCLEPDRWVKOKEEWNSIKKLTFTKKNNIOQSY 660
Qy 661 YSNINLEGEYFYMADLKDEAKKELMENIKRKKNESNLENRNYLLENALIELLDH 720
Db 661 YSNINLEGEYFYMADLKDEAKKELMENIKRKKNESNLENRNYLLENALIELLDH 720
Qy 721 LKETATICKDNNTNACETSHNATTPNCVPRGTOPTKNIKEIAOYFKRSAYEARNRG 780
Db 721 LKETATICKDNNTNACETSHNATTPNCVPRGTOPTKNIKEIAOYFKRSAYEARNRG 780
Qy 781 LHKIKGAHEGITYKRGGRKDEKXNLCIYIKHNSRNIGFSNGPCDGDGIGOTRFRV 840
Db 781 LHKIKGAHEGITYKRGGRKDEKXNLCIYIKHNSRNIGFSNGPCDGDGIGOTRFRV 840
Qy 841 GTEWEVDEHMRKDHEDYMPRRRHICTSNLEHLOTDDHPLNINYDOLVNNSEFLDVL 900
Db 841 GTEWEVDEHMRKDHEDYMPRRRHICTSNLEHLOTDDHPLNINYDOLVNNSEFLDVL 900
Qy 901 LSAYEANKIIRMYKKEKNKLGPREVTDPKHOTTCIRAIRSFADIGDITGRDLMWRNG 960
Db 901 LSAYEANKIIRMYKKEKNKLGPREVTDPKHOTTCIRAIRSFADIGDITGRDLMWRNG 960
Qy 961 DMVYLOHLETFVGNHKSILKSGKNDYNDAPRYLKLRENWMEANAKYWEAMKCDIKY 1020
Db 961 DMVYLOHLETFVGNHKSILKSGKNDYNDAPRYLKLRENWMEANAKYWEAMKCDIKY 1020
Qy 1021 LKRSKGHSQSSYCGYSDHTPLDDYTPQKLRMTTEWAEWYCYVOKKEYDLEKCKECK 1080
Db 1021 LKRSKGHSQSSYCGYSDHTPLDDYTPQKLRMTTEWAEWYCYVOKKEYDLEKCKECK 1080
Qy 1081 DKDNGOCTSESGTCTKCEACNEVNDITGLMEQNNISDYKELHEQAOQSVNSGT 1140
Db 1081 DKDNGOCTSESGTCTKCEACNEVNDITGLMEQNNISDYKELHEQAOQSVNSGT 1140
Qy 1141 EASSTAKNHIDRVIEFLSELGYOONGSKNSGSTDESAYIGNTYENYVGAFLHDTGNP 1200
Db 1141 EASSTAKNHIDRVIEFLSELGYOONGSKNSGSTDESAYIGNTYENYVGAFLHDTGNP 1200

Qy 1201 DDCQSONEFCDEKSDGKNEKAYAFRDKPODDHAGACGCKSGSPRVOLTKTKKAEKDETE 1260
Db 1201 DDCQSONEFCDEKSDGKNEKAYAFRDKPODDHAGACGCKSGSPRVOLTKTKKAEKDETE 1260
Qy 1261 CXTVNDILKENDGKQVDECHKKNKSNCTPDMQCNINLVEDPRVCMPPRQKLCVHFLA 1320
Db 1261 CXTVNDILKENDGKQVDECHKKNKSNCTPDMQCNINLVEDPRVCMPPRQKLCVHFLA 1320
Qy 1321 NDNEIKKLOSOYNLKEAFIKSAAETFFSWTYYSKQSEGNELDKELKEGIPPAFLRSM 1380
Db 1321 NDNEIKKLOSOYNLKEAFIKSAAETFFSWTYYSKQSEGNELDKELKEGIPPAFLRSM 1380
Qy 1381 FYTBDYDPLFGTIDISKGHEGSKLRQIDSLFRKNGDOKSPNGKTRQEWMTESHEIWE 1440
Db 1381 FYTBDYDPLFGTIDISKGHEGSKLRQIDSLFRKNGDOKSPNGKTRQEWMTESHEIWE 1440
Qy 1441 AMLCALVIGAKKDPFENNYGVNNVXFSDKSTLEEFKAPQFLRWLTWEYDYCYAROK 1500
Db 1441 AMLCALVIGAKKDPFENNYGVNNVXFSDKSTLEEFKAPQFLRWLTWEYDYCYAROK 1500
Qy 1501 YLKDVQEKCSNDOLKCDTECNKCEDEYVYXMKKKEMIPQDKYKPERDKRFRDQHG 1560
Db 1501 YLKDVQEKCSNDOLKCDTECNKCEDEYVYXMKKKEMIPQDKYKPERDKRFRDQHG 1560
Qy 1561 VMTDYGTNATDYLNKRFTASGDKRGSASVYOARNOLLEKQAYYDADHCCCTYFIEN 1620
Db 1561 VMTDYGTNATDYLNKRFTASGDKRGSASVYOARNOLLEKQAYYDADHCCCTYFIEN 1620
Qy 1621 DDKYTNISSKDKCKGLVKEANTGAIKQWNGPNVNNLKELTBDVLEPPSRILICHAID 1680
Db 1621 DDKYTNISSKDKCKGLVKEANTGAIKQWNGPNVNNLKELTBDVLEPPSRILICHAID 1680
Qy 1681 GNTYDPEVKNENGIRKRLMVAATTEGYNLCQYTKREKKEKIKTSDAHKSYEVPPCSAM 1740
Db 1681 GNTYDPEVKNENGIRKRLMVAATTEGYNLCQYTKREKKEKIKTSDAHKSYEVPPCSAM 1740
Qy 1741 KYSEYDLRDIIILGIDNLEDEKQTEENLKIFPNKNGTSVGKSDSTGPDGSTARKEFN 1800
Db 1741 KYSEYDLRDIIILGIDNLEDEKQTEENLKIFPNKNGTSVGKSDSTGPDGSTARKEFN 1800
Qy 1801 ENKECVNANMICYKKGRODNGSNGARSDEDLKKGCSVSPDDYPMGNRDEGTAYOPL 1860
Db 1801 ENKECVNANMICYKKGRODNGSNGARSDEDLKKGCSVSPDDYPMGNRDEGTAYOPL 1860
Qy 1861 RMEFENGEDFCRKEKELEKLVGACNDYTCGDNEDEKRRKCTACTOYKKEFISEMRPOYER 1920
Db 1861 RMEFENGEDFCRKEKELEKLVGACNDYTCGDNEDEKRRKCTACTOYKKEFISEMRPOYER 1920
Qy 1921 QIKKYEENKDKIYSEHPVAKDAEDAREYLDOLKIKCEKSSGDCYKCKMDYSTORLTDG 1980
Db 1921 QIKKYEENKDKIYSEHPVAKDAEDAREYLDOLKIKCEKSSGDCYKCKMDYSTORLTDG 1980
Qy 1981 NSQNMPSASIDDEPKVEEGCNCQVPRGPRVREPPSPRVSILSKATASKKAATAAPTK 2040
Db 1981 NSQNMPSASIDDEPKVEEGCNCQVPRGPRVREPPSPRVSILSKATASKKAATAAPTK 2040
Qy 2041 QPKKVENLTTEKRAQOTRTRRAAQOTRKTSTATTESDVGTWVKAILNKPDSRGIEGC 2100
Db 2041 QPKKVENLTTEKRAQOTRTRRAAQOTRKTSTATTESDVGTWVKAILNKPDSRGIEGC 2100
Qy 2101 NKRTYGQVPMKWCIVGSKSENEENGICMPRRRKLCLINNIQYLYNETENKRDNDIKAEAFIK 2160
Db 2101 NKRTYGQVPMKWCIVGSKSENEENGICMPRRRKLCLINNIQYLYNETENKRDNDIKAEAFIK 2160
Qy 2161 CAEIFOFLMILYIILENPAEENLONGTIPDEFRKIMYYTGVDKDMFGDIDISNDKII 2220
Db 2161 CAEIFOFLMILYIILENPAEENLONGTIPDEFRKIMYYTGVDKDMFGDIDISNDKII 2220
Qy 2221 TYTNSVTTIILNENKCKODKDEELARKIFWEKKNKFTMEGNIYGLYHLDEMEKEKIR 2280
Db 2221 TYTNSVTTIILNENKCKODKDEELARKIFWEKKNKFTMEGNIYGLYHLDEMEKEKIR 2280

QY 2281 DNYQYDMTKLPSPLEEFYKRRPOFLMWTENAEFCNKRREOLKLEACKEKECGSND 2340
DB 2281 DNYQYDMTKLPSPLEEFYKRRPOFLMWTENAEFCNKRREOLKLEACKEKECGSND 2340.
QY 2341 GTOCEACACVYQNFITKMKTEYERQOREKFKDKGKKYKDPSTERDIEKATCAHEYL 2400
DB 2341 GTOCEACACVYQNFITKMKTEYERQOREKFKDKGKKYKDPSTERDIEKATCAHEYL 2400
QY 2401 NMLKELCGNKCSCOMQKPSOLPKTTQOSQSDANDMPESLDLYPDEEFNKCCEPBLSKK 2460
DB 2401 NMLKELCGNKCSCOMQKPSOLPKTTQOSQSDANDMPESLDLYPDEEFNKCCEPBLSKK 2460
QY 2461 GSWIHTKTKTEKTPMNCVEKAAYYLSKEAENMDITLKEKPIPISTYKESKSNWTNN 2520
DB 2461 GSWIHTKTKTEKTPMNCVEKAAYYLSKEAENMDITLKEKPIPISTYKESKSNWTNN 2520
QY 2521 NPODPPKPYAPKTYIGRRNPPCENRENRKVDYEMKCYNKSRYOKKRCVPPRRHMC 2580
DB 2521 NPODPPKPYAPKTYIGRRNPPCENRENRKVDYEMKCYNKSRYOKKRCVPPRRHMC 2580
QY 2581 LRLNDEIKTERLKDENVYLLKMWRTARNIGIDILKFNPSNCGAMPICDTMYSPADLG 2640
DB 2581 LRLNDEIKTERLKDENVYLLKMWRTARNIGIDILKFNPSNCGAMPICDTMYSPADLG 2640
QY 2641 DIVRGTDMLRIGGYLPPEVETIKYVEYIYGKRNKNGKRNKYNDVQTERSAMWANDRKD 2700
DB 2641 DIVRGTDMLRIGGYLPPEVETIKYVEYIYGKRNKNGKRNKYNDVQTERSAMWANDRKD 2700
QY 2701 IMKAMTCKAPEDAKLFRKGMOGEFERTLLIODCGHKDPPVDYIIPORRMWTEASEY 2760
DB 2701 IMKAMTCKAPEDAKLFRKGMOGEFERTLLIODCGHKDPPVDYIIPORRMWTEASEY 2760
QY 2761 CRLAMELEKFKKSCDCHTSRCKNDYDENKCEQCKTRCOEYKKNVLLKWSLFDIOSNK 2820
DB 2761 CRLAMELEKFKKSCDCHTSRCKNDYDENKCEQCKTRCOEYKKNVLLKWSLFDIOSNK 2820
QY 2821 YKELYQPIYTKISTYDHOVQYQAKLTKRSECVSEFSFYETHETSKCLNTYKKNENDGSS 2880
DB 2821 YKELYQPIYTKISTYDHOVQYQAKLTKRSECVSEFSFYETHETSKCLNTYKKNENDGSS 2880
QY 2881 NITYYAFETPKSYKACSCITLPSKNPLDNCPLDOKKDGCKELQOTTFPGSKNDYDNLDN 2940
DB 2881 NITYYAFETPKSYKACSCITLPSKNPLDNCPLDOKKDGCKELQOTTFPGSKNDYDNLDN 2940
QY 2941 WNAVYLVNSSDOKGVLIPRRRHLCRPTIAYNRKGDKEILKLLTSAFSGOGLG 3000
DB 2941 WNAVYLVNSSDOKGVLIPRRRHLCRPTIAYNRKGDKEILKLLTSAFSGOGLG 3000
QY 3001 KYKSEELCFEAMKYADYSDILKGTDMMDISLSEKIKIETSDATEENRKTWENN 3060
DB 3001 KYKSEELCFEAMKYADYSDILKGTDMMDISLSEKIKIETSDATEENRKTWENN 3060
QY 3061 ROTWHAMLGKYKATSKVTLDEGMCOLPDEETNOFLRWLEWAKACCKKKVYSOLKT 3120
DB 3061 ROTWHAMLGKYKATSKVTLDEGMCOLPDEETNOFLRWLEWAKACCKKKVYSOLKT 3120
QY 3121 KCPRSNEDNEASELLROPGCONDIRKYISLNLINKNTMENLITKYLKOLKDOSSGNIDNK 3180
DB 3121 KCPRSNEDNEASELLROPGCONDIRKYISLNLINKNTMENLITKYLKOLKDOSSGNIDNK 3180
QY 3181 PSEENQSYIKSKDSOCALENDINEIVGTCKNNENNEFEVYKLLPYGLIYVEDETHKN 3240
DB 3181 PSEENQSYIKSKDSOCALENDINEIVGTCKNNENNEFEVYKLLPYGLIYVEDETHKN 3240
QY 3241 HVLGDNKEEEOYVRKALYFTPHVDSFYQAPLFTSHRAOYDPKNDILKSSIVYYS 3300
DB 3241 HVLGDNKEEEOYVRKALYFTPHVDSFYQAPLFTSHRAOYDPKNDILKSSIVYYS 3300
QY 3301 ALGLIALHFMKKKSSVDLRLTLNIPGEGYGPMTLESKNRYIPYSGPYKGYTYIYMEG 3360
DB 3301 ALGLIALHFMKKKSSVDLRLTLNIPGEGYGPMTLESKNRYIPYSGPYKGYTYIYMEG 3360
QY 3361 DTSGDEDKYMMDLSSDITSSESEYEELDINDIYVGSFYKTLIEVYLEPSKRDIPSD 3420

DB 3361 DTSGDEDKYMMDLSSDITSSESEYEELDINDIYVGSFYKTLIEVYLEPSKRDIPSD 3420
QY 3421 TPSENDPRTNRFIDDEENELKHPFVSOLPENTEPNNYKCAPIPMNTEPTWLSNDPEEK 3480
DB 3421 TPSENDPRTNRFIDDEENELKHPFVSOLPENTEPNNYKCAPIPMNTEPTWLSNDPEEK 3480
QY 3481 PFIIISHDRDLYGKEISYNNKSTNTNDIPMARNDSYRGIDLINDSLVNLILYMM 3540
DB 3481 PFIIISHDRDLYGKEISYNNKSTNTNDIPMARNDSYRGIDLINDSLVNLILYMM 3540
QY 3541 KY 3542
DB 3541 KY 3542
RESULT 2
ID 08T326 PRELIMINARY; PRT; 3287 AA.
AC 08T326;
AD 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, last annotation update)
DE PFEMPI (Fragment).
GN TM180VAR2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM180;
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy.";
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL: AJ420411; CAD20867.1; -;
FT NON_TER 3287 3287
SQ SEQUENCE 3287 AA; 383550 MW; 58F8E866FC244536 CRC64;
Query Match 66.8%; Score 12954.5; DB 5; Length 3287;
Best Local Similarity 72.7%; Pred. No. 0;
Matches 2440; Conservative 267; Mismatches 524; Indels 125; Gaps 43;
QY 12 MGAASLEBDASPIIKESHKSARNVLERYAKNIRPSYAKEHVDSLKGDITKAEFRG 71
DB 1 MGAASLEBDASPIIKESHKSARNVLERYAKNIRPSYAKEHVDSLKGDITKAEFRG 60
QY 72 GPSTPVANKHNYYPYPCNLDHKEHTNLRDYDVLNRHPCHGREGQNRFEDESECGNKR 131
DB 61 GPSTPVANKHNYYPYPCNLDHKEHTNLRDYDVLNRHPCHGREGQNRFEDESECGNKR 120
QY 132 YKRRKNAIACAPRRRRMCKNLEALNDITONTIHLGLVAVYAKREGSYNNNHK 191
DB 121 YKRRKNAIACAPRRRRMCKNLEALNYINTONIHDLGLVAVYAKREGSYNNNHK 180
QY 192 TSDACTALASFPDIDGIVGIDMFKNPVHDKVEYGLREYFKLIHDMEDVEVKNYPPDG 251
DB 181 TSDVCTALASFPDIDGIVGIDMFKNPVHDKVEYGLREYFKLIHDMEDVEVKNYPPDG 240
QY 252 SGNYIKLREAMVNNRKNKVEAITCDASYSKGYTFMOSSESTPLFSNKGCH-KOGKVP 310
DB 241 SGNYIKLREAMVNNRKNKVEAITCGALPKSAVFLQSEDKQLFLYKPCGHNNKNDLPTN 300
QY 311 LDVVPOLRPFDMGESEFCRRNITKLKVVDSORNMKERYCYSHNGHDCITTMKGIHL 370
DB 301 LDVVPOLRPFDMGESEFCRRNITKLKVVDSORNMKERYCYSHNGHDCITTMKGIHL 360
QY 371 LDNKCJDCSTKCYFVEYWLGNOCFAFKKQKEKEKEIOSLISDNKRVNNINSEYKQFY 430
DB 361 LDNKCJDCSTKCYFVEYWLGNOCFAFKKQKEKEKEIOSLISDNKRVNNINSEYKQFY 420
QY 431 EKLKEFOYATNTFTLNLNLENGKYCKGGLPGEKDIPTTNSADDKGIFYSBYCOVCPDCGV 490

```

Db 421 DQLEKHNHNLDTPLNLENEGYCKEKLKESDITFTNSSDDKGIFFRSQYCYQCPDGV 480
Qy 491 KCDIKYTHKSDNREVRNEDYKPPMGVAPUTNTVLYSGNEODLTOXLENCNSSTN 530
Db 481 KCDGTATYTHKSDNREVRNEDYKPPMGVAPUTNTVLYSGNEODLTOXLENCNSSTN 540
Qy 551 KDKNOMWECYKDEKENTNRCKLEONTEINNDPKIISFHFNFELWYLLDRTIKWMDL 610
Db 541 KDKNOMWECYKDEKENTNRCKLEONTEINNDPKIISFHFNFELWYLLDRTIKWMDL 600
Qy 611 KTCINNTTICIDECNNGCICPDWYKORKEEENSISIKLETKKKNIOQSYSTINILFEG 670
Db 601 KTCINNTTICIDECNNGCICPDWYKORKEEENSISIKLETKKKNIOQSYSTINILFEG 660
Qy 671 YFFKVMKIDKDBAKWELMENTIRKKNEFSNLENNRDYLENAIELLDLKEFTATICK 730
Db 661 YFFKVMKIDKDBAKWELMENTIRKKNEFSNLENNRDYLENAIELLDLKEFTATICK 720
Qy 731 NNTNNECETSHNATNPVAVPRGTOPTKIKETIAOYFKRSAYEARNGRLHKLKRAHE 790
Db 721 NNTNNECETSHNATNPVAVPRGTOPTKIKETIAOYFKRSAYEARNGRLHKLKRAHE 780
Qy 791 GIYKRGRRKDEKONLCRIKIKHSNRLGFSNGCDGKGIOIRFVYVGTWEDPEH 850
Db 781 GIYKRGRRKDEKONLCRIKIKHSNRLGFSNGCDGKGIOIRFVYVGTWEDPEH 840
Qy 851 MKRDEHVIIMPERRHICSTNLEHLOTDDHPLNGNIVDVLVNSFGLDYLISAKYANKI 910
Db 841 MKRDEHVIIMPERRHICSTNLEHLOTDDHPLNGNIVDVLVNSFGLDYLISAKYANKI 900
Qy 911 IRMYKKNLKGKKEVTDKROTICRAIRYSPADIGDIIRGRDLERNNGDWKLOHLE 970
Db 901 IRMYKKNLKGKKEVTDKROTICRAIRYSPADIGDIIRGRDLERNNGDWKLOHLE 960
Qy 971 YVFGNTHSLKRGNDKYNDAPKYLKLENNWEANRAVWAMKCDIYLLKDKSGHOST 1030
Db 961 YVFGNTHSLKRGNDKYNDAPKYLKLENNWEANRAVWAMKCDIYLLKDKSGHOST 1020
Qy 1031 OSSYCSYSHTEPLDYIIPKILMWTWEMAWYCKVOKEYDKLEKCKECKDNGOGCTK 1090
Db 1021 OSSYCSYSHTEPLDYIIPKILMWTWEMAWYCKVOKEYDKLEKCKECKDNGOGCTK 1080
Qy 1091 ESGTGCTEACNEYNNDIIGLKEOMNIIISDYKELHEOAMSYNSNGIESSAKNHI 1150
Db 1081 ESGTGCTEACNEYNNDIIGLKEOMNIIISDYKELHEOAMSYNSNGIESSAKNHI 1140
Qy 1151 DNVIEFISLTYOONGSKNSKSGTSDSAVIGTNTTYENVGAYLHDITGNFDQOSQNEFC 1210
Db 1141 DNVIEFISLTYOONGSKNSKSGTSDSAVIGTNTTYENVGAYLHDITGNFDQOSQNEFC 1200
Qy 1211 DEKSGKNEKAPFDPKODHDGAGCGSGSPFVVOIKTKKKAKEKTECKTYVNDILKE 1270
Db 1201 DEKSGKNEKAPFDPKODHDGAGCGSGSPFVVOIKTKKKAKEKTECKTYVNDILKE 1260
Qy 1271 NDGKQVEDCHPKKNSNGYPRMOC-----GNINIVEDPRVCPPRROKLCVHFANDNEIK 1326
Db 1261 NDGKQVEDCHPKKNSNGYPRMOC-----GNINIVEDPRVCPPRROKLCVHFANDNEIK 1316
Qy 1326 NGGTKEVCEANVTGE-YPEMECKKIONH-----KACAPPRPKLCVHFLE---L 1304
Db 1316 NGGTKEVCEANVTGE-YPEMECKKIONH-----KACAPPRPKLCVHFLE---L 1304
Qy 1387 YRDLFTGIDIKGHEGSKLEOJDSLFKNCDOKSPNKCTOEWMTESHEIWEAMCAL 1446
Db 1446 YRDLFTGIDIKGHEGSKLEOJDSLFKNCDOKSPNKCTOEWMTESHEIWEAMCAL 1436
Qy 1447 -----VKIGAKKDEFTENYGYNNVAFSDK--STLEEFKAPPOFLRWLTWYD 1493
Db 1493 -----VKIGAKKDEFTENYGYNNVAFSDK--STLEEFKAPPOFLRWLTWYD 1483
Qy 1494 YCYTROYKLYOYCKSNQOLKJDTCKNCKCEYVYVYMK-KKEMJPODKTYIDERDKK 1552
Db 1552 YCYTROYKLYOYCKSNQOLKJDTCKNCKCEYVYVYMK-KKEMJPODKTYIDERDKK 1542

```

```

Db 1480 YCHTROYKLYEVESTKSDOLKJDTCKNCKCEYVYVYMK-KKEMJPODKTYIDERDKK 1539
Qy 1539 YCHTROYKLYEVESTKSDOLKJDTCKNCKCEYVYVYMK-KKEMJPODKTYIDERDKK 1529
Db 1540 YCHTROYKLYEVESTKSDOLKJDTCKNCKCEYVYVYMK-KKEMJPODKTYIDERDKK 1530
Qy 1613 GCTKEFENDOKTNTSSKXCKGIVKAEATGAIKQONKPNYNNILKELTEVLEPPSRRL 1672
Db 1603 GCTKEFENDOKTNTSSKXCKGIVKAEATGAIKQONKPNYNNILKELTEVLEPPSRRL 1662
Qy 1673 RICEFALDGNVDPPEVKEBNGELKRLMEVAATEGYNLGOYKEKKEKKEKITSDAHKSY 1732
Db 1732 RICEFALDGNVDPPEVKEBNGELKRLMEVAATEGYNLGOYKEKKEKKEKITSDAHKSY 1722
Qy 1733 EYPPGSAKYSYDLDIILGIDNLEDEKOTENKIKIPNKGTSYVSGSDSTYGNDS 1792
Db 1792 EYPPGSAKYSYDLDIILGIDNLEDEKOTENKIKIPNKGTSYVSGSDSTYGNDS 1782
Qy 1793 TARKFEMNEKCYVNNAMICYKRGDRDNGSNGARSDEDLKKGGSVSDDYPRGKRD 1852
Db 1852 TARKFEMNEKCYVNNAMICYKRGDRDNGSNGARSDEDLKKGGSVSDDYPRGKRD 1842
Qy 1853 ESTAYOFLRMEFMEWDECKHKEKLEKLVACNDYTCGDNEDKRRKCTDCTOYKPFIS 1912
Db 1912 ESTAYOFLRMEFMEWDECKHKEKLEKLVACNDYTCGDNEDKRRKCTDCTOYKPFIS 1902
Qy 1913 EKKPOYKOIKKYGNKIKITSEHPVAADAREYLDOLKIKICEKNSGDCEYCKMKDY 1972
Db 1972 EKKPOYKOIKKYGNKIKITSEHPVAADAREYLDOLKIKICEKNSGDCEYCKMKDY 1962
Qy 1973 STORLIDNSONMPASLDDEKEVEYKNCOVPRGPRVRETPSPRSLSTAKTAKE 2032
Db 2032 STORLIDNSONMPASLDDEKEVEYKNCOVPRGPRVRETPSPRSLSTAKTAKE 2022
Qy 2033 AKTAPPTOPKVENLTTWEM--AOTTRAAQOTR-----RTSTATTEEDVGMV 2083
Db 2083 AKTAPPTOPKVENLTTWEM--AOTTRAAQOTR-----RTSTATTEEDVGMV 2073
Qy 2084 KALISNKPDSRGIEGCPKTYGOYPRMGCIYVSKENENICMPRRKLCINNOIYN 2143
Db 2143 KALISNKPDSRGIEGCPKTYGOYPRMGCIYVSKENENICMPRRKLCINNOIYN 2133
Qy 2144 YETENKRDNDIKAFICAALETOLMILKYIENPAENELONGTIDEKRIKITYYGD 2203
Db 2203 YETENKRDNDIKAFICAALETOLMILKYIENPAENELONGTIDEKRIKITYYGD 2193
Qy 2204 YKDMFPGDINSODKILTVNSVTIL--NENKCKKODK--KDELRKIFWENKKTIEWG 2261
Db 2261 YKDMFPGDINSODKILTVNSVTIL--NENKCKKODK--KDELRKIFWENKKTIEWG 2251
Qy 2262 MIVGLYTHLNDENKERTIDNYOYNDMTKLPSLSEEVKRPOLRWFTWAEFCKRKE 2321
Db 2321 MIVGLYTHLNDENKERTIDNYOYNDMTKLPSLSEEVKRPOLRWFTWAEFCKRKE 2311
Qy 2322 QILKLEAGCEYECNGSNDKTOECACAVTYONFKIMWTYEREREREKPKKDKGKYK 2381
Db 2381 QILKLEAGCEYECNGSNDKTOECACAVTYONFKIMWTYEREREREKPKKDKGKYK 2371
Qy 2382 DYESTERDIEKATCAHAYLYMMLKLELOGNDCSCMOPSSQLPKTYOQOSSANDMPS 2441
Db 2441 DYESTERDIEKATCAHAYLYMMLKLELOGNDCSCMOPSSQLPKTYOQOSSANDMPS 2431
Qy 2442 DYESTERDIEKATCAHAYLYMMLKLELOGNDCSCMOPSSQLPKTYOQOSSANDMPS 2442
Db 2442 DYESTERDIEKATCAHAYLYMMLKLELOGNDCSCMOPSSQLPKTYOQOSSANDMPS 2432
Qy 2443 LDVYVPEEF--KCBPPELSKSGSMITHKITEPKIPNVCVEKAAVYYSKEANNMOITLAE 2500
Db 2500 LDVYVPEEF--KCBPPELSKSGSMITHKITEPKIPNVCVEKAAVYYSKEANNMOITLAE 2490
Qy 2501 KFIPIESTKEKESKNWTN--NPPCOPKPYADPKYIGRRNPCEENEFNRYVDEYMKY 2558
Db 2558 KFIPIESTKEKESKNWTN--NPPCOPKPYADPKYIGRRNPCEENEFNRYVDEYMKY 2548
Qy 2559 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2618
Db 2618 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2608
Qy 2619 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2619
Db 2619 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2609
Qy 2620 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2620
Db 2620 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2610
Qy 2621 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2621
Db 2621 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2611
Qy 2622 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2622
Db 2622 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2612
Qy 2623 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2623
Db 2623 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2613
Qy 2624 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2624
Db 2624 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2614
Qy 2625 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2625
Db 2625 KNSKFOYCKRVCVPPRRHEMLRNDIEKIRLKSNTYLLMVRRTANNEIDIIKKNYN 2615

```



```

OY 2619 -SENCAMNPICDTMKYSPADIGDIIVGTDLR--IGYLPVEVEIKLYKFEYIYKWMN 2675
Db 2584 YKKNPKRRNQLCDAMKYSFADLADIIIGTDIYKPIGA--NEVENTRALIFENIYKMMT 2641
OY 2676 KKKGNKRYNDVOTFSSAMMDANKRDIWKAMTCAPEDAKLFKGRMDGERITLLIODKCG 2735
Db 2642 ENKGIKRYNDIOSFRAVMDANREYWKMTDAPDSMLFKLENGELIPNILLISQHKR 2701
OY 2736 HADDPPVDYIPQFRMMTMESEYCKALMELEKFKKSDCKTSD-RCKNDYDNKCE 2794
Db 2702 YNDHPVDYIPQRLMKMEGEYCKILNEKYNDKNDCEKCLKDKCSDDGICKOR 2761
OY 2795 OCKTRQOEYKFNVLKMKSLFDIOSNKKYKELRQPTTKISTY-----DHVONFYOK 2845
Db 2762 SCKEKCKEYTKLILHLKSGFENIQKKKYNEL-----YTKIQNNRRGFINDNDKNVIEFEK 2816
OY 2846 LKTFSECEGVSEFSEYLAHETSKCLNFKENNDGSSNIRTYAEETPKSKKACSCILPSK 2905
Db 2817 VKMI-NNCVNGTPDKYLDKAIHCIAHNFQNGIKS--KPYAANNHPEKKYKSHCSCTI-TH 2872
OY 2906 NPLDNCPTDQNDCKELOTFTFCSKNDYDNNLDNNAVLVNSSDDNKGVLLIPRRRL 2965
Db 2873 HPLDNCPTNKTAYCKTIIHNPCLITKNDNLETTGTGVVDNKKDKKGVLPVPRRL 2932
OY 2966 CRRPTAYNRYKRGDEILKLLKLLTSAFSGOLLGOKYKSEELCEPAMKYSYADYSIIR 3025
Db 2933 CYRKLITGNRYIRNEKDNILKNDLSAFSHGMLGKTFNDYNOGMSMKYSFADYDIIR 2992
OY 3026 GTDMDT-----LSEKIKIFETSN-----EATENRKTWENNRRQIWMALCGY-KIAT 3075
Db 2993 GTDMLGGSIIDFNKDKMPFENNSENIGKTTISREOWMEKKKRYNAMALCGYOKGRK 3052
OY 3076 SYVTLDEGCOPLKDEETNOFLWLEMAKQACEKKHYSDSLKTCPSRNEDEFEASEL 3135
Db 3053 NGEWDKNNKNCNPTEDGTQFLRWLEMAQACVKNHVRDLSKTCRCSEKDNFEASEL 3112
OY 3136 LKPGCGQNDIRKYISLILIKTMENTLITKYOLKQDSSGNDINKRSEENVOSYIKSKDS 3195
Db 3113 LKPGCGQNDIRKYISLILIKSMENLITKYOLKQDPSGNDINKRSEENVOSYIKSKDS 3172
OY 3196 OCALRENDINELVGTGKNNNEFKVLEKLYPGLVFNDETHKKNVLDNGIKKEEQYTR 3255
Db 3173 ECDLENDINELVGTGKNNNEFK-VLKKLYPGLVFNDETHKKNVLDNGIKKEEQYTR 3231
OY 3256 PALATFEPTHVSFOAPLESTHRVAYQPKNDILKSSISVYVSLGLIALHEFM 3311
Db 3232 PALATFEPTHVSFOARLEPLTYREKYPKNDILKSSISVYVSLGLIALHEFM 3287

```

```

RESULT 3
OBT325 PRELIMINARY: PRT: 1615 AA.
ID OBT325:
AC OBT325:
DB 01-JUN-2002 (TREMblrel. 21, Created)
DB 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DB 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE PTEMP1 (Fragment).
GN TM284VAR3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN NCBI_TaxId=5833;
RP SEQUENCE FROM N.A.
RC STRAIN-TM284;
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyse S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy.
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420412; CAD20868.1; -
FT NON_TER 1615 1615
SQ SEQUENCE 1615 AA; 188639 MW; FCA896C00866DEA6 CRC64;

```

```

Query Match 38.0%; Score 7367; DB 5; Length 1615;
Best Local Similarity 83.0%; Pred. No. 0;
Matches 1358; Conservative 69; Mismatches 169; Indels 40; Gaps 12;

OY 12 MGNAAASLEGDAKSPILKSHKSARVNLRYAKNIRHPSKYAKEHVDLSKGLDTKAEFRG 71
Db 1 MNTESLSLEGARSPIESSENSPRVNLRYAKNIRHPSKYAKEHVDLSKGLDTKAEFRG 60
OY 72 GSTPKNKNNYYPYPCNLDHKEHTLRYDVALRHPCGRGRNRRDEDEESCGKIKIN 131
Db 61 GSTPKNKNNYYPYPCNLDHKEHTLRYDVALRHPCGRGRNRRDEDEESCGKIKIN 120
OY 132 YKKNDAICAPRRRRHMDCKNLALNDINTQNIHDLGNVLTARYESIYNNPHRG 191
Db 121 YKKNDAICAPRRRRHMDCKNLALNDINTQNIHDLGNVLTARYESIYNNPHRG 180
OY 192 TSDACTALARSFADIGDIYRGIDMFKPNYHDKVEYGLREVFKKINDGMEVAKNDYNDPG 251
Db 181 TSDVCTALARSFADIGDIYRGIDMFKPNYHDKVEYGLREVFKKINDGMEVAKNDYNDPG 240
OY 252 SGNYKLRBAMNVRNKNKYWEALITCDASYSGVFMOSESNTPLFSNPKGKHGKYP-TN 310
Db 241 SGNYKLRBAMNVRNKNKYWEALITCGALPKRSATFMOSEDKOLFSTPKGHNKDKPLTN 300
OY 311 LDYVPOYLKRFDEWGEFRCRKNRIKLKYKDSGRNDEKRLYCSHNGDCTTITMKGILH 370
Db 301 LDYVPOYLKRFDEWGEFRCRKNRIKLKYKDSGRNDEKRLYCSHNGDCTTITMKGILH 360
OY 371 LDKCTDCTCKKVFYEWVLGNQOBAKROKYEKEIOLSYLSDNKFVNNINSEYKQY 430
Db 361 LDKCTDCTCKKVFYEWVLGNQOBAKROKYEKEIOLSYLSDNKFVNNINSEYKQY 420
OY 431 EKLKTOYATNDFLNLWEGKCKGGLPGEKOTITNSADGDGIRYSEYCOVCPDCGY 490
Db 421 DQLRDKNKNDLDTFLNLWEGKCKGGLPGEKOTITNSADGDGIRYSEYCOVCPDCGY 480
OY 491 KCDGIRYTHKSDNDRBVNNEDEKPPWGVKPTNITVLYSGNEGDIYOKLENFCNSTYV 550
Db 481 KCDGIRYTHKSDNDRBVNNEDEKPPWGVKPTNITVLYSGNEGDIYOKLENFCNSTYV 540
OY 551 KKKNNKMECYKYDENINCKLEQNTNINNDNKIISFNHFLWATYLLRDTIKANDKL 610
Db 541 KKKNNKMECYKYDENINCKLEQNTNINNDNKIISFNHFLWATYLLRDTIKANDKL 600
OY 611 KTCINNTTHCIDEGRNCLCPDRWYKOKKEEENSITKLTFTKKNYQOYSYNNILPFG 670
Db 601 KTCINNTTHCIDEGRNCLCPDRWYKOKKEEENSITKLTFTKKNYQOYSYNNILPFG 660
OY 671 YFEKYNKLDKDEAKWKELENIKRRKNFESNLNNRQVLENAIELDLHLKETATICKD 730
Db 661 YFEKYNKLDKDEAKWKELENIKRRKNFESNLNNRQVLENAIELDLHLKETATICKD 720
OY 731 NNTNEACETSHNATTPCYKPRGCTOPTKNIKIAYOFKRSAYEERNGLHAKGAHE 790
Db 721 NNTNEAYETSHNATTPCYKPRGCTOPTKNIKIAYOFKRSAYEERNGLHAKGAHE 780
OY 791 GIYKRGRRKDFDNCRTIMIKSNRNLGFSNCPGCKGSDGIQYRFVYGTMEVDPDEH 850
Db 781 GIYKRGRRKDFDNCRTIMIKSNRNLGFSNCPGCKGSDGIQYRFVYGTMEVDPDEH 840
OY 851 MKRDHEDVIMPPRRRICTSNLEHLOTDPHPLNGNITVDLVLNNSFLGADVLSAKYERANKI 910
Db 841 MKRDHEDVIMPPRRRICTSNLEHLOTDPHPLNGNITVDLVLNNSFLGADVLSAKYERANKI 900
OY 911 IRNYKRNKNNLKGPEYTDKQHTTICRAIRYSFADIGDIIRGDLWERNGDVYKLOGHLE 970
Db 901 IRNYKRNKNNLKGPEYTDKQHTTICRAIRYSFADIGDIIRGDLWERNGDVYKLOGHLE 960
OY 971 TYFGNTHKSLKGGNDYNDAPKYLLKRENWMEANRAKAYEAMKCDIKYLRKDSGHOST 1030
Db 961 TYFGNTHKSLKGGNDYNDAPKYLLKRENWMEANRAKAYEAMKCDIKYLRKDSGHOST 1020

```


QY 1013 AMKDCIKYLKDSHOSYSGYSDHPTPLDYIPQIKLMTMTEAEMVCKYQKKEPYDL 1072
 Db 709 ---C-----KSHDEDEHNEYC-----CDLIPKSVDDO 732
 QY 1073 KEKCEKCKDKNDQOGCTKESGTGCTKCTEACNEYNIDILGKMEQWNIISDKYKELEHOAO 1132
 Db 733 EEDDEEV-----DEEKE-----744
 QY 1133 MSVNSNGIEASSTAKNHDNRNVIETLSELYOONGKSNKSGTSDASAVIGTNTYENGA 1192
 Db 745 -----SSQTTKRNI-----SOKGTGTSACVACAIYGV-L 777
 QY 1193 YLHDITGNPDCCOSONECEKSDGKNEKYPADRKQDDHDGACGCKSGSKPTVQJKTAK 1252
 Db 778 OOKSNGSINDCNKAKNR-----793
 QY 1253 KAEKOTECKTVNDILKENDKKQVEDCHPKKNSNGYPMOGCINLVE--DPRVCMPPRR 1311
 Db 794 -----KKN-----EMQCKNTFTVDSGCVCMPPRR 818
 QY 1312 QKLCVHFLANDNEIKKLOQOVNLKEAFISAAETFPFSWYYSKSDGEGNELDKLEKEG 1371
 Db 819 KSJCINHLLEBQTK---NKYOLREAFIKCAKETMLMDKYANDKNEAEEL---LKKG 872
 QY 1372 IPPAFIRSFYTFYTDYRDLFTGTDISKHGBGSKLEQJDSLFKNGDQKSPNGKTIQEW 1431
 Db 873 IPDEFKRIEYFEGDEPDECLENDMGK---DVDKVKKNINKYFNNSKRGFKKIDPENW 929
 QY 1432 TRESHEIWEAMCALYKIGAK-----KDDPTENYGYNNYKFSDK-----STLLEFAKRP 1481
 Db 930 NENGPINNMICALIHADTKDSIRKND---NYKIEKTYILAKRGSGNMTLSEAKRP 985
 QY 1482 QFLRWLMEYDYCYTRQYKLYADVOERKSN--OLKCDTECNKCEYVYKRRKKR-EM 1538
 Db 986 KFLMFEVEMYDYCKEROXYLFEVASTCKSIDGQLKCDRGCKNNKDEYKRYRKKKEEM 1045
 QY 1539 IPQDKYKDERKRRDRDHIGVMTDYGTNATDYLNKFTASCDKPSA-----SVY 1593
 Db 1046 NLODKYKDKREKNGIDKPIGTILYVLANKEYLKKFKFASCTYSSKKAONSATEEV 1105
 QY 1594 ORNIQILEKQAYYADADKHCCTEFTIENDKYTNISSKDKCKGLVYKANTGAIQWOKNBP 1653
 Db 1106 KNIETLSEBOYDADQYGCFTFI--HDKYISKISGRSNCCGLNSAKKNIKIMRSDEK 1164
 QY 1654 NYNNLKE--LTEDVLFPSRRLICFHALDGNVTDPEVKDENGILKRLKMEVAATEGYNLO 1711
 Db 1165 DYAFLLKRRNSGCVFPPSRRLICFHALDGNVTDPEVKDENGILKRLKMEVAATEGYNLO 1224
 QY 1712 YKKEKEKEKIKTSDAHXYSYEPPPSAMKYSFYDLROIILGIDNLEDEKQKTEEMUKI 1771
 Db 1225 YKKEKEKEKEATEEAHXYSYEVOQPSAMKYSFYDLROIILGIDNLEDEKQKTEEMUKI 1284
 QY 1772 FKNKGTSGVKGSDSTGNGSTARKFEMNENKCYVNNAMICGYKRRDQNSGNSARSDE 1831
 Db 1285 F-----KSESNGSGOROTFFANNKKGCVWEAKKCGIKHGRDQNSANSASSDQ 1332
 QY 1832 DLKKGCSVSDDDYPMGKNRDEBGTAYOFLRWEFMEWEDFCFKHKELEKLVGACNDYTCG 1891
 Db 1333 DLKKGCSVSDDDYPMGKNRDEBGTAYOFLRWEFMEWEDFCFKHKELEKLVGACNDYTCG 1392
 QY 1892 DNEDDRKCKTDACTOYKKTPISEMKPOYEOIKYKGENKIKYSEHVAADADEARAYTLK 1951
 Db 1393 DNEDDRKCKTDACTOYKKTPISEMKPOYEOIKYKGENKIKYSEHVAADADEARAYTLK 1452
 QY 1952 QLKKEICENSGDEYKCMKDVSTQRLTDSNOMSPASLDEPEKEVGKCNQVPRPRV 2011
 Db 1453 QLOKSC--NSGKCD--CMNKKSF-----SNGNMNPASLDEPSTYDRCCGQPPPPP-- 1502
 QY 2012 RRTTBPRLSLSKATASKKEAT-----APPTKOPKK-----2044
 Db 1503 ---PAPARPAPARESGNRYGRSEPEDEGPLPLPPLPPKPPKGGGAGRILSTPRNG 1559
 QY 2045 --VENLTERAQTTRBRAAQOTRRKRTSTATTESVGVNVAKILSNKPPDSRGIT--EGCN 2101

Db 1560 TIEDEDEDEGEKAEALASEETSK--EVEQKEDTTEKCYIVANILTGKONLDACN 1617
 QY 2102 PKTYGQYPR---WGCIV-----GSKENENNGICMPRRKKLK 2135
 Db 1618 QK-YG-YQ-RHMKWCKVTPTTSSSTSERGASNNKNNLDSTKSDKNGSCICIPRRKRLX 1675
 QY 2136 INNIO-YLNYETE-----NKRNDIK-----EAFICAIETQFLM 2171
 Db 1676 IKKIQEMASGNQAGNGTSGDSTGASNPQNGVSTSPQVALLHAFVESAATEFELMD 1735
 QY 2172 KY-----IENPAE-----NELQGTIPDEKRTIMYTYGYDKMFG 2210
 Db 1736 RYKKEKEIKQOQETGLVASSETSEDETHPQNKLGSTIPLDLRQMFYTLGDRICVG 1795
 QY 2211 TDISNKKIITVNSYTTLLENKKKOD--KKD-----EELKLFEMENKKEFIEGM 2263
 Db 1796 ---KTPDGLDITYASADKNDMDKIQAKIQOILPKKDIPTPSYKTPQOTWNNKHAESINGM 1852
 QY 2264 YGLTYHLTDENEKIR-----DNYQNDMTKL-----2291
 Db 1853 YALTYR--TTPSGEKPRQIPEVTKLLEDEKGTQSNKYQYKV--KLEEBETSGAKPKSTE 1910
 QY 2292 --TPS-----LEEFYRPOFLRFTMAEFCCKKKEQLKLEAGC-----KEYEONG-- 2337
 Db 1911 SSSPSEMTPLDIFISRPFRFLYEMGQNFCEKRRKRLDIDISNCLKDGDKOYSGDGA 1970
 QY 2338 --SN-----DGKTOEAEACVYQNF IKMKTEYEROREKPKKDKGKYYK--DY 2383
 Db 1971 GSNIDVNRKIDIFDLGPRCAKPCSSRYKWKIKKQYEOEAYVQOKEQCKKEFNH 2030
 QY 2384 PSTERDIEKATCAHEYLNMKLELGNKDCSCQKPSOLPKTTQOSSDANDMPESID 2443
 Db 2031 NGFRVILGCTTAPGDL--QTLKNGPC-----KSENGDHBE-- 2065
 QY 2444 YPPEENKCEPCLSKSGSIHTKITEPIPNVCYKAYISKRAENMDTLLEKPI 2503
 Db 2066 --DEINFSO--PDVYTRPA--TNCPTCPKFKYNC-----2093
 QY 2504 PIESTKEESKNSMTNNNDPPKPYAPDKYIGRRNPCEENREBRKVDYEMKCYKNSKF 2563
 Db 2094 -----KNGNCADJNGKNGKTP-----2111
 QY 2564 YOEKKRVCPPRREHMLNLEDEIKTERLDSYLLKMYRRANRNGIDILKFNSENGC 2623
 Db 2112 -----IDAQNEBOMQOTAKEFVMIVSDKSTNGFEV-----NDLNEC 2147
 QY 2624 AMNPIDCTMKYSFADJGDIYRGDMLRIGYLPPELKLKYVEEYLYGKRRNKKGRNXY 2683
 Db 2148 A-----GAD-----IFQGI-----2156
 QY 2684 NDVQTPRSAMWANDNRKDIWKM-----TCKAPE-----DAKLFKRGMDGFERITLLIQDKC 2734
 Db 2157 -----KENINSGRNVGLDYCK--PEKYNQDKVNRK--ENDG-----2189
 QY 2735 GHKDDPPVDYIPORFRMTWSEYCYKALMEELFKPKSCHDKTSRCKNDYDENKCE 2794
 Db 2190 -----YIIQIALLRRAVENF---LEDYKKIKHKSISCHTSTEEK-----KSTCD 2232
 QY 2795 QCKTRCOEYKNFYKKSLEDDIOSNRYKELYEDPIYTKISTYHVNQNFVQKLTFFSECS 2854
 Db 2233 --CGKCK--CGQWIKLKEEMKKIKENY-----VEENNSE-- 2265
 QY 2855 VESFSEYLETSCKLYKFNENDSSNIRTYAFETPKSYKEACSTLPSKNLDNCPTD 2914
 Db 2266 --DFSQNLNLFLETLITQIPVADVQGVNITLSNFDPP-----CGSADANS-----2310
 QY 2915 QNRDCKELOTFFGCKNDYDNNLDMNAYLYVNSSDNKGVILIPRRRLCTRPITAYN 2974
 Db 2311 QKRDG-----NENAIQDCMIN-----RLOOKANCTEOPSGSKOCTTPTPTTL- 2352
 QY 2975 YKRGDEILLKKLLTSAFSGQLGQKYKSEELCEFAKMYSAVDSDIIGTDMDTSL 3034

Db 2353 -----EDDETFD-----DDI----- 2362
QY 3035 SEKIKKIFETSNEATENRKTWMENNRHQAHLGCKIATSKVTLDEGCOLPKDEEN 3094
Db 2363 -----ETDN-----PVANHPCISGVITOTETVLEEKCAVPABEEK 2399
QY 3095 QFLRLIEMAKOKCKEKHVSLSLTKCPRSNEDNEASELLAPQCONDIRKYLISLNL 3154
Db 2400 D-----EKKREDP----- 2408
QY 3155 IKNTMENLNIKYOLODOSSGNIDN-----PSEENVOSYIKSKDSOCALINDINEIVNG 3210
Db 2409 -----AEBOGAIKPGSPGAPQPPSPS----- 2431
QY 3211 TKNNENNEKEVLKLYLPGLYFEDETHKNHVLGNKEEQTVREKALYFTPHVDSFY 3270
Db 2432 -----DTEENHV-----TPREDDPP 2447
QY 3271 QAPLSTHVAQ-----YDRKNDILKSSIVYVAGLILALHEKKKFKSSVDLL 3321
Db 2448 PAD--DTRPPSPPLPPADQPPDP--TIIQTIIPGVALALGSIAPFFMKKKTKHPVDLF 2503
QY 3322 RLINIPQEGMPTLESKNRYIPYRSGPYKGTUYIEMEGDTSGEDKXVWMDLSSSDITSS 3381
Db 2504 SYINTEPKGDYDIPTLKSKNRIPYRSGSKGTIYMEDS--DEDKTAFMSDITDITSS 2561
QY 3382 ESEYEELDINDIYVPGSPYKTLIEVLEPSKR--DIP--SDDTSPND--TPRTNRFIDEM 3437
Db 2562 ESEYEELDINDIYVPGSPYKTLIEVLEPSKNGINIPHAGEPLDDMVPITNTFTDEM 2621
QY 3438 NELKHEFVSQYLPNTEPNK--NYK--SADIEMTEPTLXSDPEKPFITSHDRLTYG 3494
Db 2622 NELKHEFVSQYLPNTEPNK--NYK--SADIEMTEPTLXSDPEKPFITSHDRLTYG 3494
QY 3495 KEISTININKSTNNDIPNANRDSYRGI 3523
Db 2681 EISTININK--VNNDIPMSDKNGTYGTI 2706
RESULT 5
Q26032 PRELIMINARY; PRT; 3006 AA.
ID 026032
AC 026032
DT 01-NOV-1996 (TREMUREL 01, Created)
DT 01-NOV-1996 (TREMUREL 01, last sequence update)
DE 01-DEC-2001 (TREMUREL 19, last annotation update)
GN Variant-specific surface protein.
Plasmodium falciparum.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5683;
RN 11
RP SROUENCE FROM N.A.
RC STRAIN-FCR3;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.2., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A.,
Peterson D.S., Ravetch J.A., Wellem T.E.;
RT "The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium falciparum-
infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL; LA0609; AAA75397.1;
DR InterPro; IPR004258; PFEMP.
DR Pfam; PF03011; PFEMP.2.
SQ SEQUENCE 3006 AA; 343769 MW; 8FD5FP475F943C74 CRC64;
Query Match 21.9%; Score 4254; DB 5; Length 3006;
Best Local Similarity 30.7%; Pred. No. 2.3e-197;
Matches 1133; Conservative 483; Mismatches 1002; Indels 1076; Gaps 133;
QY 31 SHKARNVLERYAKNIH-PSKYAKHEVDSLKGLDTAFERGGSTFVNKNHNYPPCN 89
Db 8 TTKTAKEVLDIGKEVOEKATEDALTYRNDLQNLGSAKPHG--VPIDVKN-----PCD 59

QY 90 LDHKEITNLRYPDVLNRHPCHEGREQNREDESESCGNKIRNYKKNDAIACAPRRRH 149
Db 60 LVEIHTNH-VGRRKENPCGREERFESDVLGQCAKNTIKDSVTNSGACAPRRH 118
QY 150 CDKNLEALNDINTONIHDLGNVLTAKYEGESYNNPH-----KGTSDACTAALARSAD 205
Db 119 GDRNELIKTDENTSTHDLVDLTAHGESLVKHKEYIKKNRNPNTCTVLARSAD 178
QY 206 IGDIVYGDIMF-----KPNVHKVETGLREYFKKIHDM-----EDEKKNYN 248
Db 179 IGDIVYGDILYGDYDEEKEKRLKELNKKFQGIHDDVAKTSGRNTGKSAEAKRYN 238
QY 249 PDGSGNYYKLEPAMNVRNRYWEAITCDASYKSYFQSSNTPPLSPNPGCHKQKVP 308
Db 239 -DATGNYTKLRDMWANNARNDVWKAITCDADNDEYFENSSDGLYVNSGCGRNEKVP 297
QY 309 TNLDPVQYLRWFDEMGSEFCRKRNIKLKFKDSCRNDEKERYLCSHGNGHDOCTTIWKGI 368
Db 298 TNLDPVQYLRWFDEMGSEFCRKRNIKLKFKDSCRNDEKERYLCSHGNGHDOCTTIWKGI 368
QY 369 LHLNKCITDCTCKVEFWLGNQOEAQKOKERYEKEIOSYLSNDKFNVINSEYKQ 428
Db 358 CSSDNCCTACSNICLAYDAMLRNORNEERKOKIKYKIEKFKSSPSKSNISNKTYNE 417
QY 429 FYEKLETOYATNDFLNLNEGKYCKGLDEGDIPTFNADKGIIFYSEYQVCPDC 488
Db 418 FYENNGKRYETLQNLFLKLNKGMTCOEKEEVEVDF--NKDDMYFHREYCOCPDC 475
QY 489 GVKCDGILYTKHSNDRER-----VNNEKYKPPWGVKPTNITVLSNGNOGDIOTKLE 541
Db 476 VVOCCKGKCTEDKRNKCRSKIKIILQSE-----PLEIHVLSNDKOGDITKLE 527
QY 542 NFCSSTNYKDKNNOKWECYKDEINRCKLEONTETLNNDPKTISHNPFELAVTYLR 601
Db 528 VFGSSTNYESRNVQKWKCKNKNNDYNNCEMNTSYSDSTANVMSVECFHSAKMLLI 587
QY 602 DTKRNDKLTCTINN--TFHIDECNRCCLCFDEWVYKKEEENSIKLTKKKNIDQSY 660
Db 588 DTIKWEHQKNCINNTVTCESKICNCEYEMIKRHEHEKAVNAGNNRRSYIY 647
QY 661 YSNINNFEGTFEYVMDLKDE--AKKKEIMENIKRRKNEFSNLENNRDYLENAIELLD 719
Db 648 YNNISRYVDSFLFQVMAALDODEGKWDPTEDLK--KKEFSPKTNPTGKSQAIEFLD 706
QY 720 HKEFATICKDNNTNEACETSHNATNPCKVPRGGQOPTNINIEIAQYFRSAVEARNR 779
Db 707 HKNDAITCKDNNSNESCDSKVKTNPCGNPSASNNLVYRRLAEMQRYARKOLEKR 766
QY 780 GLH-KLKGAKHEGTYKGRKDFKDLIRIMIKHS--NRNLGFS--NGPCDGK--GTGDS 833
Db 767 GGEINLKGDATKGYRQGGPADGK--NVCISINQNHNTVQNNNAAYYQGFCTGKDSNG 825
QY 834 IOTRFYVGTMEVDEPEMKRKHEDVYMPRRRICTSNLEHLOTDPHL--NGNIYDVLN 892
Db 826 V--RMKGTSPWK--PGQIDMSADDIYMPRRRHCHCSNLEYLOTGKPGKQD--GKLVN 880
QY 893 NSFGLVLSAKYEANKIIMYKKEKNLKGKPEVTPKQDTICRAIRYSPFDIGIING 952
Db 881 NSFGLVLSAKKADKADKIIETLYKKNN--KSVLTDPEDNESACRLARSFADLGIING 937
QY 953 RDLMEKNGMVKLQGHLETFYVGNHSLGKGNKDYNDAPKYLKRENMWEANRAKVE 1012
Db 938 RDLMDKNSDAKRLQTLNKEIFTKIEELPEDIKKYDKDGTGHLKRLRENMWEANRQVNR 997
QY 1013 AMKCDIKYLRKDGSHOSTOSYSGYSDHTPLDITPQKIRMTWAEWYCKQYKYEYKL 1072
Db 998 AMKCALENDKDK-----CNGIPIDEDYIPQRLRMTEVAEWCKQKQSRLYNLT 1045
QY 1073 KECKECKKDNQOGCTKESGTCTCTACNEVNDIILAMEONNITSDKKEKLEHGO 1132
Db 1046 VADCKSCCKK--AKSTQKDG--DCTFKCAACDNVKNKIKPMEQOEKIKNKTAQLKAL 1102

QY 1133 MYSNSGIEASSAKNHNIDRNVEFLSELYQON-GKSNKSGTSDSAVIGTITTEBNG 1191
 1103 DSVNGEESKSKKTAQADQVHFLAELTRKSGGKGNVKTIVSPTTPTLLYSAA 1162
 QY 1192 AYI-HBTGNFDDGOSONEFCEKSDGKDNKVFPRDKPODHGACGKS-GSPTRJOIK 1249
 1163 GYHHELGKRVGNCNTOKEFYTSK-NKTAFKDPKGYEACCKDNBNP-Q 1212
 QY 1250 TKRAAEDECTEYVDILKENDGKROVEDCHP-KNSNGYEDMOCGINLVEDPRVCM 1307
 1213 PAKKDEDA-CDVVKPLKDKGETDDIOGCNKYKAGKDKYGMOCNSOIHHTHNGACM 1271
 QY 1308 PPRROKLCYHFLANDNEIKKLOSOVNLKEAFIKSAAEFFSYYKSKDGBENDELKEL 1367
 1272 PPRROKLCYHFLANDNEIKKLOSOVNLKEAFIKSAAEFFSYYKSKDGBENDELKEL 1367
 QY 1368 KEBKIPAPLRFMSYFEGYRDLPFGDLSKGBESKLEOJDSLF-KNGDOK-S 1421
 1326 KGNINIEGFRKOWYTFGDRDLPFGHJST-HAYISGVPKYITLLEKENDAKYAKON 1384
 Db 1422 PNKTRQEWMTESHEIWEAMCALV-KIG-AKKDTEENYGNVNYKFSK-STLLEEF 1477
 1385 SNNEELDDMDQKIDMEGMLCALFHKISDEEKKKEIKKYSYKLNESPKGSNKVEF 1444
 QY 1478 AKRPOFLRWLEWYDYCYTROYKYLKDYOKC-KSND-OLKCDTECNKCEYDYK 1531
 1445 AKRPOFLRWLEWYDYCYTROYKYLKDYOKC-KSND-OLKCDTECNKCEYDYK 1531
 QY 1532 MKKKKEMIPDKDYKDERDKRRROHIGVAVDYTG-TNADYLNKRFASCGD 1585
 1505 IYKKV-EYTK-QGKFDKEL-IDKBYEBSFTKDSAEYIKK-C- 1546
 QY 1586 KPGSASVYQNIOLLEKQAYYADKHGCTKTEENDKYTNISSKCKGLVKEANTGAI 1645
 1547 -LDDTCNCKQKYNKNTYNT- 1566
 QY 1646 KMNKGPNNNNKELTEDVLFPSRLRICFHALDGYTDPREKENDGLKRLMEVAATE 1705
 1567 -PNKYTN-SNLEKCE-QOPPO-EPPEEGAR-SDS 1600
 QY 1706 GYNLGOYKKEKKEKIKTSDAKHYSYEVPCSAKMYSPFDLDIILGIDNLEDEKOTE 1765
 1601 G-PROTPRAGSDARSNTYPPRPAGDTIVAEV-OEBEED 1645
 QY 1766 ENLKIFNKGTSVGKSDSTGNPGSTARKFFWENNEKCVNMAKICGYKRGDDGNSN 1825
 1646 -DGLPE 1651
 Db 1826 SARSDIEDLKCGSPDDDYPMGKNDEGTAQOPLRWFAEM-GEIDCKHKELEKLYGA 1884
 1652 D-ODEVEVAGAEBEED-LDVGA-BILGRITNSPDEDEEASEE- 1694
 QY 1885 CNDYCGDNDEKRRKCTDAQYKFKISEWKPOYKOKIKYGENKDKITSEHPVAKDAED 1944
 1695 -DDDDDDADQPTLEVYQ-GEETAEHDHOD 1721
 QY 1945 AREYLDKOLKICEBNSGCEYKCMADVSTORLTDGNSOMPASLDEPEKEBCKCQY 2004
 1722 TTEETVQ-EKAEEDKDGGE-TPOKET-QPVEVYPCD- 1757
 QY 2005 PRGPRVRARPTSPRVSLSKATASKKEAKTAPPTOKPKEVNLTTMRAOTRRRAAO 2064
 1758 -IVATLTTEETLEACPTKYNRE- 1782
 QY 2065 TRKRTSTATTESDVTGMVAKILSNKPSDSRGIEGCPKPYGOKPKGCIVGSKENENG 2124
 1783 -KFPNNKCI-SSGSDASGS 1799
 QY 2125 ICMPPRRKILCINNIOYLT-ETENKRDND-IKEAFICAALETOLFMLKY-IE 2176
 1800 ICILPPRRKILYLIHIEGVDTVSSDGETTPIITHDALREAFIOTAAVEFELMHRYKIK 1859
 QY 2177 NPAENELONGT-IDEFKRLMYTYGYDKDMFEG 2210

Db 1860 EKEROEELONGTLEPPAQKVSPEDNPEHPOKLIKEGKIEPEKROMFYTLGDRDLCVG 1919
 QY 2211 T-DI-SNDKITVTNSVTLINENN-KKODKKDEELRKIE 2250
 1920 VKDDVQALAEASDNKSGDKNIDISEKIKSVLEKGEOTPECPKQOTTTKEE-W 1975
 QY 2251 WEKRRPFWEGMITYGLYHLD-ENEEKIRDN-YOYNDMTKLTPS- 2294
 1976 WQKNGEIHMAICALI-HHTDTRQYDDQYKGLFENGKATPRNSQOYKYNVITISSVSN 2034
 QY 2295 -LEEVKRPQFLRWLEWYDYCYTROYKYLKDYOKC-KSND-OLKCDTECNKCEYDYK 2341
 2035 GGPIGNIKLEOFASRPFLFWLEWYDYCYTROYKYLKDYOKC-KSND-OLKCDTECNKCEYDYK 2341
 QY 2342 -KTECAACVLYONFIKKKTEYEROR-EKPRKDKGK-YKOD 2383
 2095 EKPCKDGSFEYTLKCPSCASCSYKWKISRKDEDEFKQAYEKOKDABGNNDPKER 2154
 QY 2384 PSTERIEKATCAHEYLNMKLELGNKDCSCMOKPSOLPKTTOQOSSDANDMBESLD 2443
 2155 SKTLRNVNDAAA-FLN-SLK-NGPCS-KNDDSVODEIK 2189
 QY 2444 YVPEE-FNKCEBELSKGSMITHKITEPKIIPMNCVKAAYLSKEANMDITLKE 2500
 2190 FDERKTFGEHYCKPCKSK-ITVCK 2213
 QY 2501 KFIPIESTKESKNSVTNNPCDPKPYADKYIGRNPENREERFKVDEMYCKYN 2560
 2214 K-ENHCDNSKP-NDC- 2227
 QY 2561 SKFYOEKRVCPVPRRHHMLRLDEIKIERLDSNYLKMVBRTANEGIDITKFNSE 2620
 2228 -RINSISAE-DIEKRSNT 2245
 QY 2621 NGCAMPID-IMKYSFADGIVGCTDMLRIGYLPPEILKYVEFYIYKMKRNRK 2678
 2246 QDVYMS-VSDSNTNGKNEFDLNCICA-GIFKGI- 2277
 QY 2679 GRNRYNDVQTFRSAMMDANRDKIMKAMTAKAPADAKLFRKRGMDFERITL-10DKC 2734
 2278 -REDYWK-GEICVDICTLEKTNNERY 2304
 QY 2735 GHRDDPEVDYIPORF-RWMTESBYCKALMELEKFKSCDCKTSRCKNDYDEN 2791
 2305 SAKENDNKNQIILIRVLFKRWL-ESFLEDNKINDKISHC-MKND- 2347
 QY 2792 KCEQCKTROEYKFNFLYKMSLFDIOGNKKELYEOPITYKISTYDHVONFVKLTKFS 2851
 2348 KSPJCINGCCKNCVCKWLEKKSSEWKVREY-INQY- 2385
 QY 2852 ECSVESFSEYLETSCLANKFENNDSSNIRYAFEEETPSKYKACSCITLPSKNPLDC 2911
 2386 -RDKNSN-EAPF-VKSLFELTILPQIP-V 2409
 QY 2912 PTDQNK-DGCKELQTFPFGSKNDYDNLDWMAVLYLNSDDNKGYLIPRRRRLCTRP 2970
 2410 VTDKGRKHSITOLKRLKCSSEKSEN-SNEKDYV-LC- 2445
 QY 2971 TAYNVRKGDKEILLKRLLSAFSGOLGOKYSEBELCEAKKYSADVSIDYIKGDM 3030
 2446 -LTK- 2449
 QY 3031 DTSLEKIKRIFETSNEATENRKTWNENRQIWMALCGYKATSKVTLEDEGMCOLPKD 3090
 2450 -LEDAKAN-C-KD 2459
 QY 3091 EETNOFLRWLEWYDYCYTROYKYLKDYOKC-KSND-OLKCDTECNKCEYDYK 3150
 2460 QASGE-PCQUTSENDDDEDLLE- 2483
 QY 3151 LNLIKTMENTLNIKKYKOLKDSGNDKPKSENNOSYIKSDSOCALENDINEIYVG 3210

Db 2484 -----ENVEAPNT-----CPKVEPEPVV--EBEKC-----DLAEPASK 2516

QY 3211 TKNNENFEKFKLVKLYGLVFEVDETHKNHVLVDGNIKKEEOQVRAKALFFPHVDSFY 3270

Db 2517 ESSFEENSEG-----GSMSEONPKRSPKEEPPPPPSSETDP----- 2553

QY 3271 QAPFSTHRAQVD-PKN---DIKSSIVYVSAALGLALHFMKKKKSSVDLRLTNI 3326

Db 2554 -PPAPPTIOPSQADQPTNSISDLSITIPGIALALSIYFLFKKTKSSVDLRLTNI 2612

QY 3327 POGEGAPLTESKRRYIPYSSGPCKCTYIYMEGDTISGDEDXKMMDLSSDITSSSEYE 3386

Db 2613 PKGEGYPTLKSSNRKIPYASDKRKGITYMEGDS--DSGHYED--TDTVTSSEYE 2668

QY 3387 ELDINDIYVPGSPKRYKTLIEVLEPSKRDIPSD--DTPS-----NDPRTNRFTIDEM 3437

Db 2669 ELDINDIYVPGSPKRYKTLIEVLEPSKRDIPSD--DTPS-----NDPRTNRFTIDEM 2728

QY 3438 NELKHDFVSQYLPNTEN---NNYKSDIPMTEPNTLYSDNPEKPFITSHDRDLYTG 3494

Db 2729 NTLKDEFISQYLOSEPNDVNDYTSQNSSTNTNITTTSRHNYEERKPFIMSHDRMLYTG 2788

QY 3495 KEISYNNSTNTNNDIPNARNDSYRGIDILINDSL 3530

Db 2789 EEINYNVNM-VNTMDPIPRDNNVYSGIDLINDAL 2823

RESULT 6

Q26030 PRELIMINARY: PRT: 3026 AA.

AC 026030: 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE Variant surface protein (Fragment).

GN VAR.

OS Plasmodium falci-parum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE OF 2044-2922 FROM N.A.

RC STRAIN=IT 4/25/5;

RX MEDLINE=95330813; PubMed=7606788;

RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A., Peterson D.S., Ravetch J.A., Wellens T.R.,

RT "The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falci-parum-infected erythrocytes."

RL Cell 82:89-100(1995).

RN [2]

RC SEQUENCE OF 2044-2922 FROM N.A.

RA STRAIN=IT 4/25/5;

RX Hudson-Taylor D.E., Peterson D.S., Pinches R., Newbold C.I., Miller L.H.;

RT "Switches in the expression of Plasmodium falci-parum var genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes."

RL Cell 0:0-0(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=IT 4/25/5;

RX MEDLINE=99094502;

RA Smith J.D., Kyes S., Craig A.G., Fagan T., Hudson-Taylor D., Miller L.H., Baruch D.I., Newbold C.I.;

RT "Analysis of adhesive domains from the A4VAR Plasmodium falci-parum erythrocyte membrane protein-1 identifies a CD36 binding domain."

RL Mol. Biochem. Parasitol. 97:133-148(1998).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=IT 4/25/5;

RA Smith J.;

RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RC STRAIN=IT 4/25/5;

RA Kyes S., Smith J.;

RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL: L42244; AAD03351.1;

DR InterPro: IPR004258; PFEMP.

DR Pfam: PF03011; PFEMP. 2.

FT NON_TER 3026

SQ SEQUENCE 3026 AA; 348124 MW; EC0A971A82193788 CRC64;

Query Match 20.7%; Score 4010; DB 5; Length 3026;

Best Local Similarity 29.8%; Pred. No. 1.5e-165;

Matches 1095; Conservative 480; Mismatches 1151; Indels 954; Gaps 127;

QY 30 ESHKSAENVLYRYANNI-RHPSRYAKEHVDSLKGLDFAEFGSPSTVNNHNYYPYC 88

Db 13 EDDCKAKHVLDSIGKRYKVKVANYSSQKGLSLNAIFENEPKGGQTEND-----PC 67

QY 89 NLDHREHNLKRDVNLNHPCHGREGQNFDEDESECGN-KIRNYKRNDAIACAPRRR 147

Db 68 KLYEYHTNVTKGH-GREHPCKRKTGKRPDVGGECDNRKIKD--SKNMGACAPYRRL 124

QY 148 HMCDEKLEALNDINTQNTIDLLGNLYTAKYEGESIVNNHHR-----GTSDACTAL 199

Db 125 HLCVARNLENISALDKINDTLLADVYCLALHEGOSTIDYKYQAQYASSFSFSQICTML 184

QY 200 ARSFADIDIVRGIDMEKPNVHDK--VETGLREYFKIHGDMED-EYKNDYNPDGSGNY 256

Db 185 ARSFADIDIDIRGKDLVIGNKKERKLEDKLTKIFGIYKELTPRAKDHKKDPDNNF 244

QY 257 KLRSEAMVNNKYNWEALITCDASYSKGFMSSENTPLFSNPKGHHQK-----VPT 309

Db 245 QLRSEAMVNNKYNWEALITCDASYSKGFMSSENTPLFSNPKGHHQK-----VPT 304

QY 310 NLDYVPOYLRFDEMGEEFCKRRNIKLYKVDSCR---NDKERLYCSHHGDDTTTWK 365

Db 305 YEDYVPOYLRFDEMGEEFCKRRNIKLYKVDSCR---NDKERLYCSHHGDDTTTWK 363

QY 366 KGLHLNCKCDDCTCKCVFEMVLCNOEAFKKOKEYEKEIOSYLSNDKVV---NNT 421

Db 364 KHKLMDNACGCFSCSDRFRKIAKODEFKKNNKYTBK---KNDTJTITTEGTI 420

QY 422 NSEYKOFYEKLKTOYAINDTFLNLNEGRKCGK---GLPGEKDTTFNSADKGIY 477

Db 421 NNMTRKDFYKHLER-KKTYDAFLNLNKECKNHPEVEGKKTYIDFDNIE---TFS 476

QY 478 RSEYCOYCPDGCYKCDGKITTHKSDNDRVNN---EDYKPPGVKPTNITLYSG 530

Db 477 HTEYREPCPMCGIE-----EKGDKMKRINDHSACKEEELTPKENAKYTKINVLTSG 529

QY 531 NEQDITQKLENCNSSTNYKDKNN-----QKMECYKDE----- 565

Db 530 EGHEDIKRLKEKCTKQNGGGSDCGGSSDLCPEWQCYDPLDERVGGGEVDKIK 589

QY 566 -----NINCKLEQNTENNDPKIISFNFEFLWYTLDRITKMNKLKTCIN-NTT 618

Db 590 GAGGACIFERKMGKRYK-----KQTFNFPNFAVHLAKSIDMRQLTKCSLSD 643

QY 619 THGIDECNRNCLCPDRVYKKEEENSIKLFTYKKN-IQOYSYNNINLFEGYEKKVND 677

Db 644 KKECKGKSGCECFKWKIEKREKWKVDOFNKQDFLEWKHYLVLETLLEYVRENIO 703

QY 678 KLDKDEAKMELMENIKRKKNEFSNLENNNDYLENAITELLHLKFTATIC-----KDNK 732

Db 704 KAGGDLKSIQEMKMKIKENKQNNKTKD-----EDLADVLFDEKEADCDJLHEDDD 759

QY 733 TNAACETSHNATNCPVKGPGTPTKNIKELTAQYKRSAYEARNR--GLHKLGAHE 790

Db 760 DDECVETIEKIPNNPC-----SGTRHRAVKNVAADYRAAPOOLNRRBAGKRTLAADSO 815

QY 791 GIYKRGRRKDFKDNLCRLMIKHSNNLGFNSGPDGCKGTGIGIOTRFVYGEVDEPEH 850

Db 816 GHYNGKANSVLKD-VCDITTOYNSA-IGDSKDPCKGKDG-----FKIGTPWITVVK 867
Qy 851 MKRDEHYIMPRRRHICSTNLEHOTDHPNGNIVDVLVNNSEFGVLLSAYEANKI 910
Db 868 KTSYADVFLPPREHMCJSLNLENDVDVTVNNV-----VANKFLVYLLSANKAEWI 922
Qy 911 IMKYK-----KNLKGCEKVTDRKHOTTICRAIRYSPADIGDIIIRGDMERNGDWYKLOG 967
Db 923 KQRYEPNQNHNHKKR-----CRALKSPFADIGDIIIRGDMERNGDWYKLOG 970
Qy 968 HLEIYVGNHLSLKGKNGKYNDDAPKYLKRENNWEANRAMEKCDIYLNKSKSH 1027
Db 971 NLVTFJGKATVOKRGIDISTYNTDGNHQLKREDMWEANRAMEKCDIYLNKSKSH 1022
Qy 1028 OSTOSSYCGSDHTPLDDYIPQRLMTEMAMWYCKVQKREYDKLEKCEKCKDNNGOG 1087
Db 1023 -GKINGC---ATPYDDYIPQRLMTEMAMWYCKVQKREYDKLEKCEKCKDNNGOG 1072
Qy 1088 CKRESTGCTKTEACNEENNDIIGLKEBWNIIISDKY--KELHEQOMSVNSGIEASST 1145
Db 1073 ICKNWKDCDAKCTEACKREYKRIQPKDMWEKLELEYALSYH----- 1115
Qy 1146 ANKHIDR-----WIEFLSELYOQ-NGKSNKSGTSDSASVIGNTYENVGAY 1193
Db 1116 AKNDSTRMAFGGIDPDYQOVVHFKELOALIKSSTSKRRKSTDAITPTPTPYSTAAGY 1175
Qy 1194 LHDYGNFDCQSONEFCDEKSDGKNEKYAFBDRPODHAGCCKSGSPTRVQITKKR 1253
Db 1176 IHQEIGNACQJOIKHRCNDNDK---KYVFERKPRDHEACCTENVE-----KPK 1222
Qy 1254 ABEKOTE---CKTYNDILKENGKOVECHCKKNSGTPD---WQC---GNINLEVPD 1303
Db 1223 KEEDIDKLVCAIVKALTATTODLTKA---COOK---YGHPRHMGKCISETTKSSDSG 1277
Qy 1304 RYCMPPRROKLVHFLAN-----DNKIKLOSOMVLEKA 1337
Db 1278 SLICVPRRRKLVVPLTKAEBATEBPTSPAGGEITLTPPATASQAQKGBSL--LLRA 1335
Qy 1338 FTKSAAEFTFFSWYKSKDGBENLDELKELKEGRIIPAFRLSMFYTPGDRFLFTG 1394
Db 1336 FIOSAVERFEFLMHKKY--MDNNGGDAEDLKEGRIIPAFRLSMFYTPGDRFLFTG 1394
Qy 1395 -DISGHEGSKLKE---QIDSLFKNGDOKSPNG-----KTQDMWTEHSHL 1438
Db 1395 VIKALEASSDNKSGNNIKESIDKIEAILKQSGSLGGLVTPPNVKNKRTWMDQNAKH 1454
Qy 1439 WEAMLCALV---KIGAK-KODFTEN-----YGVNNVKS 1468
Db 1455 WHGMCVALYKEDTGAKKGTSTODPTAVKGLMDGKPKEDKDYKIVKISSVPSN 1514
Qy 1469 --DKSTLEEFAPKPOFLRMLEWMDYCYTRQKYLKVOECK-----SNDOL 1515
Db 1515 VPSGDTRKEEFRRPFEMLEWMDYCYTRQKYLKVOECK-----SNDOL 1515
Qy 1516 KOD-----TECNKCEDEYKYM-KKKKEMIPQKYYKDERDOKRPRDRIH 1559
Db 1575 ICKTDTSRNNFTIDHCPCKLEKCIKRYMIEKKEEFHNOKNNEKEFN--DLKKE 1631
Qy 1560 GWAVDTGTNATDYLNRKFTASCGD-KPGSASVQVRN-IOLLEKQAYDADKHC-GCTK 1616
Db 1632 G-----YSSFN-----NPLASLNCHKGEHNRDKNKIEFNHNTKTPSPSTACAPY 1679
Qy 1617 FLENDKYTNISSKCKQGLVYEANTGAIRWONKGNPNYNNLEKTEDELVLFPSRLRICE 1676
Db 1680 YGVKCKR-----KNGCEETIHKTDLNG-----QNDNNTYDILKVLID 1716
Qy 1677 HALDGWYTPDYVDENGSLKRLMEVAATEGYNIGQYKKEKKE-----KITSDAKH 1729
Db 1717 --RKGSSNDEELKNNVNNSTSLFKDSSV-----QYWKCKKNEVDQCIIDFLDIDIK 1767
Qy 1730 YSEYVPCSAKYSFYDLIDIIIGIDNLEDEKQTEENLKPIFNKNGTSGVGSDDSTG 1789
Db 1768 Y-----MEFNVPFQRLRIFVHDYINLKKIKPCK-----TDEKKNKCIINGC 1811

Qy 1790 POSTAKFEFNNENKCECVNMMIIGYKRGDRDNGS-----NSASDEDLKCKGCV 1839
Db 1812 KGLTECVKFKMLKQKODEMKNKIDHYEKNSLYGYIPHWKSYFVQLVFERDYKKAQV 1871
Qy 1840 PSDDDYPMKNDDEGTATYQFLRMFAWGEDFCKHKKELKLVGACNDYTCGDNEDKRRK 1899
Db 1872 IED-----ENBRKKTWCGTDCVCECNEETREK 1899
Qy 1900 CTDACTQYKFTISEMPOYEKOIKKYGKNDKRYSEHPAKXADEARLXOLKICEN 1959
Db 1900 -DEITNLKIKLOEKIESCOTQHPNG--KYRDEIIPHSD-----ETLDEDTDTTDD 1950
Qy 1960 KSGDCEYK-----CMKDVSTORLJDNQSONMPASLDEDEFEYEGKCNQVPRGPRVRE 2014
Db 1951 DMSDKIYDRKPPPCPRVDKET-----EKPK-----VLGPR----- 1984
Qy 2015 TSPRSLSLSKATSKKAKTAPPTQPKVENVLTTEMAQOTRRAAOTRKRSTATT 2074
Db 1985 ----- 1984
Qy 2075 TESDVGTMYKALISNKPDSRGIEGCPKTY---GOYPMKCIYVSKRENEGICMPRR 2131
Db 1985 -DACEIYGEIL-NGODGTKEKEBCNTKYTPKNDYPCGMC-TDKVINREBSGCPRR 2039
Qy 2132 KILCINNIOYLVYETENKRDNDIKAEAFICAALETQFLMKY-----IEN 2177
Db 2040 OKLCIHNLBHL---SEKATELRLKAFICAALETQFLMKY-----IEN 2177
Qy 2178 P-AENELONGTIPDEFKRIYTYDYDYKDMFGTIDSK-KIITVTSVTTILNENK 2235
Db 2097 PDDPKRLLEGITPEPKFOMETYGDRDLEGTIDSGHGESALGKRIISLFRNGQ 2156
Qy 2236 KRODKKDELRKIFWEKKNKFIWEGMIGLYTLDENEKERIRNY-OYNDMTKLTPS 2294
Db 2157 KSPSGKTPPE---WANDGPDIMKMGVCSLHINKGN-KEDLRKNDLNNKTYTISK 2211
Qy 2295 LEEFYKPOFLMFTWAEBCNKRREOLIKLEAGKEYECNGSNDGKTOEACAVTYO 2354
Db 2212 LEDFASRPOFLRMFIEMQGFCHERYVAKINOLKTGNEVEYEGSGQENKREACKNACEAVK 2271
Qy 2355 NFKKKTETEYERREKFKKDKGKKKDYDPTROLEKATCAHEYLMMKLKELCGKDS 2414
Db 2272 SWLDMKDDYEDQTAFFDKDKDKF--DGTSAVDVAASVHEYLLEELKLNCTGDA 2330
Qy 2415 CMQKPSQPLKTTQOSSDANDMPESLDYVPEEF--NKCEPFLSKSGSMIHTKITEPK 2473
Db 2331 CMKPSAQ---DETELLGNGTFFPAMDYPREIEGRCK-----AIPS 2371
Qy 2474 IPNVCYKAAVYLSKAEENNDITLKEKFLPIESTKEKSKSNWTNNPCDPKPYAPDK 2533
Db 2372 EPMSCVEOIAKHLREKAEKRVKYEES--LKGTAPAK-SKNDCTRID--EALKGNGSK 2424
Qy 2534 YIGR-----RNPCKRRE---NRFKYDEMYKANSFYQKRVKVCPPREHMCN 2583
Db 2425 LNKSLIDTSPASNCQSESDATDRUKIGKOWP---NKINGTEYLVPPRRKDCEND 2481
Qy 2584 LDEIKTERKDSNYLYKAVRTARNGIDIIKFNSENGCANPICDTJKYSFADGDIY 2643
Db 2482 LKNIOFNEVODSLSLEKIOHAKNBSIDILKTLNODQONAPSEIDAKYASDADGDI 2541
Qy 2644 RGTDMRLIGGYLPVEIKLYVEEYLYGKWRNKNKGRNRYNDVOTFRSAMWANDRDIK 2703
Db 2542 RGRSKID-----PTNNKTEKELOKIFKQIODNASTLSK-ELPELREKWMANDRREVN 2595
Qy 2704 AMTCAPADEAKLERGRMGFERITLID-----KCGHDDPPVDYIYQRFWMTEKSE 2758
Db 2596 AMTCVAPNDHAKLKKKNPNKNSQIIASOTEGTKRSHOSEPPDYIYDERRHFLEWSE 2655
Qy 2759 YYKALAMELEKFKKSCDHC-KTSDRCNDYDENEKCEQKTRCOEYKNFVLLKWSLFID 2817
Db 2656 YYKALAKEKNDKNCSCIKSGALCEKEDEKCEKCNCKEYKNTYVDKQSGSEFDO 2715


```

QY 2818 SNKKELEYEPIYTKISTDHYON--FVOKLTKFKSECS- VESFSEYLETSCMLNYKEN 2874
Db 2716 NOYKRLYODRTHGSPSTARNPISIEFTOKL---EDSCNDPYGADRYLCTHCTDYKTS 2772
QY 2875 ENDGSSNIRTYAFETPKSKYKKEACSCGLPSKNPLDNCPTPDONKCKELOTFTFCKNDY 2934
Db 2773 ETD--SRESYATSPYKPKDYKCKCKV-----NTPt-----SND- 2806
QY 2935 DNNLDMNMAVLYLNVSSDDNKGVLIPPRRRLCTPITAYNRYKDKKLLKLLTSAFSQ 2994
Db 2807 -----PK-----S 2809
QY 2995 GOLLGOKYKSEELCEFAKITYADYSDIITKGTMDADTSLSEKIKIETTSNEATEENRT 3054
Db 2810 PSLTGPSP----- 2817
QY 3055 WVENNRQIWMHMLCGYKIATSKVTLDEGMCOLPKDETNOFLRWLIEWAKQACKKKHV 3114
Db 2818 -----FLPK----- 2821
QY 3115 SDSLTKCPRSNEDNEEASELLRQPCQNDIRKYISLNTLINTMENLNIKYKOLKQSS 3174
Db 2822 ---KPK-----MKFYPRIGIGVLAHPFINM----- 2842
QY 3175 GNIDNKPSEENVSYIKSKDSQCALEINDINELIYGTCKNNENREKVKLLYPLGYEVE 3234
Db 2843 -----VADPTIHEVAKTFENNAVQF----- 2864
QY 3235 DETHKNHVLGNIKKEEQTVPKALYFTPHVDSFYQAPLFTSHHVAQYDPKNDILKSI 3294
Db 2865 -----HINPKTDVAP-----TKNt-----LNEVLPSKI 2889
QY 3295 SVYIVSAIGLILHMKKKKFKSSVDLRLNIPQEGYGMPTLESNNRYIPYSGPYKGR 3353
Db 2890 PYGIALALASIAFLKLKTKTHPVLDFSVINIPKSDYDIPKLSNRYIPYSGPYKGRNG 2949
QY 3354 FYIYME--GDTSGDEDKYMDLSSDIT--SSESEYEELINDIYVGPSPKYTLEVYLE 3410
Db 2950 FTTLKEIYGTDSGYDHY-----SDITSSSEYEELINDIYVHVLNITKTLLEVYLE 3003
QY 3411 PSKR-----DTPSD--DTPSN 3424
Db 3004 PSKLSGNTIPTSGKNTPSD 3023

RESULT 7
Q26031 PRELIMINARY; PRT; 3078 AA.
Q26031:
01-NOV-1996 (TREMblrel. 01, Created)
01-NOV-1996 (TREMblrel. 01, Last sequence update)
01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR-1.
OS Plasmidium falciparum (isolate Dd2).
CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2;
EX MEDLINE=95330813; PubMed=7605788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
  Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
  cytoadherence and antigenic variation of Plasmidium falciparum-
  infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL; LA0608; AAA75396.1;
DR InterPro; IPR004258; PFEMP.
DR Pfam; PF03011; PFEMP. 2.
SQ SEQUENCE 3078 AA; 349297 MW; C8037C2B03CCD7C3 CRC64;

```

Query Match 18.0%; Score 3489; DB 5; Length 3078;

```

Best local similarity 26.7%; Pred. No. 2.7e-160;
Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

QY 9 IIRGNMASSLEGGAKSPLIKESHKSARNVLYRYAKNIRHPSK-YAKHVDLSKGLTKA 67
Db 4 LAKMGPKEAAGDDIED-----ESAKHMFDRICKDYDYKVEAKKRGGLGRSLSEA 56
QY 68 EFRGSPSTPVNKHNYYYPCNLDHKEHTNLRDYDVNI-RHPCHGHEONRPEDEBSEC-G 126
Db 57 KEFKNESDPQTPED-----PCDLHKKYHTNV-----TTNINPCADRSDFRFEDEGGCOTH 108
QY 127 NKIRYKRRNDALACAPRRRHRMCKNLEALNDINTQVNIHLLGLVLTAYEESSTVNN 186
Db 109 NRIKDSQGDKNKGACAPRRLHVCQNLQEOLEPIKITHTHLLVDYCAAAAFEGOSTIÖD 168
QY 187 HP-HKGT-----SDACTALASPADIGDIVRGIDMFKNVHD-----KVETGLREVERKI 235
Db 169 YPKYQATYGDSPSOICTMILARSFADIGDIVRGIDLYLGNPOEIKORQOLENNLTIFGKI 228
QY 236 HD---GMEDEVKNYNPDSGNYKLRREAMNVRNRYKWEALITDASIKSYFMQSESNT 292
Db 229 YKLNGLAARAGND--PE---PDLREDWNTANRETVWKAITGNA-WGNTYF-HATCNR 280
QY 293 PLFSNPKGHRQKVPNTLDVYPOLRWFDMEGGEFCRKNRIKIKYKDSQR---NDKE 348
Db 281 GERTKGYCRNDQDVPYFYFDYVQPLRWFEEMADDFCKKKKKIKDYKRNCRGDKEDKD 340
QY 349 RLYCSHNGDCTITIMKGIHLNCKCTDCSTKCKVEVWLGNQOAFKKOKERYEIKQ 408
Db 341 R-YCSRNIGYDEKTRALGIRYKQICISLYACNPVYDWINNOKQOFDOKKKYDEIK 399
QY 409 SY-----LSNDNKFVNINSE-YKQPEYKIKETQATNTPTNLNMEGYC----- 454
Db 400 KYENGASGSRQKRDAGGTTNTNDYGEKKKYDLANKSEKTYDKELEKISNEICLYK 459
QY 455 --KGLGEKDIITNTNSA-----DKGIFRYSEYQVCPDGVCKDGIKYTH-KSDND 504
Db 460 DEEGGIDTFKVNNSDSTSGASGTNVESQGTFRSKYQCPQPCGYK---KVNNGSSNE 515
QY 505 RERVNN-----EDYKPPWGVKPTNITVLYSGNQGQITQKLEFCN-----SS 547
Db 516 WEKNNNGKCKSGKIYEPKPKREGITITILSKGKHDIIEKLNKFCDEKNGDTINSGGSG 575
QY 548 TNYKDKNN-----QKWECTYKDNINRCKLEQNTIENDNPK-----II----- 586
Db 576 TGGSGGNSGRQELYEEMKC-YKGEDVYKVGHDDEDEYENKMGGLCILANKKKE 634
QY 587 -----SFHNFELMYTYLLDITKANDKLTCTI-NNTTHG-IDECNRNC 629
Db 635 EGGNTSEKPEDEIÖKTINPFYFYVWAMHLKDSIHKMKKILÖRCLÖNGNRIRKCGNNKCNDC 694
QY 630 LCEDRWYKQKEEEMNSIKKLFTRKKNIQ-----OSYSNINLEPEGY--- 672
Db 695 ECFRMTÖKKKDEMGKLVQHF-KTÖNIKGGGSDNTAELIPRHDIYLYÖYNLÖEFLKGD 753
QY 673 -----FKVVDKLDKDEAKKELMENIKRKNESLNENRD-----YLENATELL 718
Db 754 SEDASEEKSENSIDAEBA-----EELKHLRETIIESDNNQASVGGVTEQGNIMDKLL 807
QY 719 DILKETATIC-----KDNNTNEACETSHNNTTNCVPRGTO-PTYNIKEIQ 766
Db 808 NYEKDEADLCLEHDEEKEKEGGDNECEENRYNCSSESQSKRRPYLVANKVAQ 867
QY 767 YFKRSAYEARNGLKHLKKAHEGILYKRGRRKDEKMLCRIMIKHSNRLNCFSGNPGD 826
Db 868 MHHKAKTQLASRAGRSALRGDISLAQFKNGRNSITLKQICKINENYSDNSGNSGPGCT 927
QY 827 GKGTGCGIÖRFYVGEWEYDEPHMKKDEHDVIMPRRHICTSNLEHLÖTDDHPLNGNI 886
Db 928 GK-DGDHGVFMRIÖTEWNSIEGKQÖTSYKNVFLPRRBMCTSNLENLDV-----GSV 980
QY 887 V-DDLVNNSFLGDVLLSAKYEANKIIRMKKKNLKGPEYDTP---KQÖTTCAIRIS 942

```


Db 981 TKNDKASHSLGDVOLAKTDAEIIKRYKDNNTI---QUTPDIQKQOEAACRAVRIS 1036
 QY 943 FADIGIIRGRDLMERNGDVYKLOHLETVENGIIHSLKG-KANDXYND--APRYLKL 998
 Db 1037 FADIGIIRGRDLMERNGDVYKLOHLETVENGIIHSLKG-KANDXYND--APRYLKL 998
 QY 999 REMWMEANRAKWEAMKCOIKYILKDSGHOSSTOSVGYSDHPLDDYIPQKIRWMEWA 1058
 Db 1097 RADMEANRHOVRAMK-----ATIGIIC--PGMPVDYIIPQIRLWMEWA 1141
 QY 1059 EWCYQKKEEDYKLEKCEKCKDNDGOGCTKESGCTKCTEACNEENDIILKMEOWN 1118
 Db 1142 EWCYQKKEEDYKLEKCEKCKDNDGOGCTKESGCTKCTEACNEENDIILKMEOWN 1118
 QY 1119 IISDKYKLEHQAOMVSNSGIEASTAKNHD---RWTEIPELYOONGK---SNK 1171
 Db 1200 KISDKNLYLQAKTSTNG---RTVLGDDDDPYOQWDFLPIHKSIAARVLYKRA 1255
 QY 1172 SGSDSAVIGTWTYENVAYLHDTGNFDDCOSONEFCDEK-----SGCKNEKYAPR 1225
 Db 1256 AGSPTEIAAAPTTPYSTAGYIHOEIGYGGCEOTQFCCKHGAFTSTYTKENKEYTFK 1315
 QY 1226 DKQODHGAOGCKSGSKPTVQJTKKKAEBEJDECTVNDILKENDGKROVEDCHPRKN 1285
 Db 1316 QPPEYATACDCINRSOTE---EPKKKEENVSACKIYKLEKNGRTYVECHPRKES 1371
 QY 1286 SNGYPMQC-GNINLVEDPRVCPMPRRQKLCVHFLANDNEIKKLOSOVNLKFAIKSAA 1344
 Db 1372 ---YPMDCKNNDISID-GACMPPRQKCLTYIAHESOTENIKEDNLDKAFIKTAA 1427
 QY 1345 EPEFWSYYSK-DGSENEIDKELKSGKTPPALRSMEFYFGYRDLFTDLSKGBEG 1403
 Db 1428 EYLSQYKSKNDSEKILDR---GLIPQFLRSMTYFGYRDLCLNTDLSKGBEG 1483
 QY 1404 SKLEOISLFGKNDKSPNGKTRQEWTEHSEIWEAMLCALVIGAKKD--FTEHY 1460
 Db 1484 AKKADIKGFKFSKDSGSPGSLRQEWKTKNGEIMKGMICALTKYVTDNDNRKIKNDY 1543
 QY 1461 GYNNVFSK-STTLEBFAKPOFLMFLMEYDYCYTRQYKLVKDOEKCS-NDOLKCD 1518
 Db 1544 SYDKVNOSONGNSPLEFAKPOFLMFLMEYGEFCAERKKEENIKDACEINISTQOCH 1603
 QY 1519 T---EENKCEDEYVYKMK-KKEMIPQDKYKNOERDKRRDRQHIGWMTDYGTANTDY 1574
 Db 1604 DAKHRCQORAOEYVENKKEFSGOTNFWLKVANQOPPEYKGYEYD----- 1654
 QY 1575 LNRKFTASCGDKPSASVYQARNIQLLEKQAYYADKHCCTKFIENDKTYNISKDKC 1634
 Db 1655 ---GVOP---LOGNEYLLQK---CQNNK-CSC----- 1676
 QY 1635 GLVKEANTGAIKWONGPNPNYNNLKELTEDVLEPSSRLRICFHALDGNVTDPEVKDENG 1694
 Db 1677 ---MDGNVLSVSPKEX----- 1689
 QY 1695 RKLMEVAATEGYNLGOYKKEKKEKIKTSDAHKYSYEPSPSAAKYSYDLRDLIIGI 1754
 Db 1690 ---PFGY---AHKYP----- 1699
 QY 1755 DNEDEKQKTEENIKKIFNKNGTSVGKSDSTGNPGSTARPFMMENKECVNNAICGY 1814
 Db 1700 --- 1699
 QY 1815 KGRDNGNSNSARSDDLKKGSGVPSDDYPMGKNRDEGTAYOFLMFAWGEDEFCCKH 1874
 Db 1700 --- 1699
 QY 1875 EKELEKLYGACNDYTCGDNDRKCKCTDACTQYKFISEMKPOYEKOIKYGENKDIYS 1934
 Db 1700 ---EKC----- 1702
 QY 1935 EHFVADAEDAREYLDKOLAKICENKSGDEYKCMKDVSTQRLTNGNSQMPASLDEPK 1994
 Db 1703 ---DC----- 1704

QY 1995 EWEKNCQVPRGPPRVRRRETPSPRVLISKATASKKEAKTADPTQPKVENULTTEMA 2054
 Db 1705 -YOGK---HVSISP----- 1723
 QY 2055 QTRBRAAQOTRRKRTSATTTESDVGMYKAILSNKPDNSGIEGONPKTYGOY---PKWG 2112
 Db 1724 ---PEAPVYVCSIVTKLR---DTNINSDAGLK-YGKTADSSWK 1764
 QY 2113 CI-----VGSKENENGICMPRRRKLCLINNIQYLANET-----ENKRD 2151
 Db 1765 CIPSDPKSGAGATGKSGSGSICIPRRRRRLYVGLQ---EWATALPQEGASAPHSRA 1822
 QY 2152 NDIKEAFKCALETQFLMYLIIIE-NPAENE-----LONGT 2188
 Db 1823 DDLRNFIOGSALETFLMRYKEEKPOGDSQOALSOLTSTYSDEDEPPDKLQNGK 1882
 QY 2189 IPDEFRIWYTYGDKYKDMFEGDISDKIITVNSVITLLENKKKDKKDEELK 2248
 Db 1883 IPPDFRLMFTYLDYRDILVHGNTSDSGNTGNSNNNTVLASGKEDMOKIOEKIEQ 1942
 QY 2249 I-----FWEKNNKFYMEGMIYGLY-----HLDNEKEKIR 2280
 Db 1943 ILPRNGTPLYKSSAQTPOKMMNEHAESIMKMICALITYTEKNPDTSAKDENKIEKD 2002
 QY 2281 DNT-----QYN-DMTRL-----TPSLEFVKRP 2302
 Db 2003 EYEFKFGSTADHGTASTPTGTQYDYEKYLEDTSAGKTPSASDPTLSDFLRP 2062
 QY 2303 QFLRMTFAEERCKRQDLKLEAGCKEY-----CN---GSN 2339
 Db 2063 PYFVLEWGNQCNKRKHIAQIKHECKVEENGSGSRGRTROYSGDEACENMLPKN 2122
 QY 2340 DGKTOE---CAEACTYONFINKMTEYERERERKFKRD---GKKYKDYPTERD 2390
 Db 2123 DGVPDLKESCAKPCSSYKMWLESKGEPEKEKAYEQKDKCVNGSKND- 2174
 QY 2391 EKATCAHEYLNMKLEKLGKDCSCMOKPSSOLPKTQOSSDANDMPESLDYVPEEFN 2450
 Db 2175 ---NGFCETL-----TTSKADFLTKG- 2195
 QY 2451 KCBCPELSKSGSMI-----HTKKITBP-KIPMNCVKAAYILSKENMDITLKE 2500
 Db 2196 --PKPNNVEGKITFPDDKTFKHTKD-CDPLKFSVVC----- 2230
 QY 2501 KFIPIESTKESKNSWTNNPCDPKPKYAPDKYIGRNPENRENRKFDYEMKCYKN 2560
 Db 2231 ---KKECDNS-----KGTDCRN-----KN 2247
 QY 2561 SKFYOEKKRVCPRRREHMLRLNDELKIRLDSNTYLMVRRJARNEGIDIIKKNFS- 2619
 Db 2248 S-----IDATDIENGVSYLEKRVASDSKS-----GFNGD 2278
 QY 2620 --ENGCAIMPICDMKYSFADLDIVRGITMLAIGLIPVLEIKLVPEYIYKGRNKN 2677
 Db 2279 GLENAC-----RGAG-----IFGCI----- 2293
 QY 2678 KGRNRYNDVQFRSAMDARKDIKRA-----MTCKAPADAKLFRGRADGERITLIQ 2731
 Db 2294 ---RKDEMKCRNVCYGVCK-PENNVAEAKG----- 2321
 QY 2732 DKCGHDDPPVDYIPQFRFMTWSEYCYKALMELEKFKKSCDHCTSDRCKNDYDEN 2791
 Db 2322 ---HIQIRALVRWEYF---FEDYNKIKH-----KISHRIKN---G 2355
 QY 2792 KCEQCKTROBYKPNFYKMSLFDIOSNKKKELYEOPITYIKISTYDHQVQFVKLTFS 2851
 Db 2356 EISPC-----IKNCVEKW---VDQKREKKEKTEER----- 2385
 QY 2852 EGSVESESEYIHTSKLANKFNENGGSSNIRTYAEFEPRK-----SYKRA 2897
 Db 2386 Q-----YK-NDNSDDDNVRSFLETLIPQITDANAANKVYILSKFGNS 2426

QY	2898	GCCTTSPKRPNDLNCPTDQNKRCCELOTFEFGSKNDYDNNLDWMNAYLVLNSDDNKVL	2957
Db	2427	CCGSASA-----NENKNG-----	2440
QY	2958	IPPRRHLCPTPTAYNVRKGDKEILKLLTSAFSOGOLGQYKSEBELCFEAMKYSY	3017
Db	2441	-----	2440
QY	3018	ADYSDILKGTMDMOTSLSEKIKKIFETSNATENRKTWMENNRQIWMHMLGKYIALSK	3077
Db	2441	-EYKAL--DOMLKLKDKT-----	2457
QY	3078	VTLDEGWCQQLPRDEFTNOFLMWLTEMKQACEKKHVSLSLTKC-----PRSNEDNEA	3132
Db	2458	-----GEC-----EKNHQS-DTECSDPQPTLED-----	2483
QY	3133	SELLQPPCCQNDIKRYSLNLINKNEMINITYKOLKQOSSGNDIKDPSEENVOYSYKS	3192
Db	2484	-ETL-----DDITEERKKNMKKICEN-LKTAQOEDB-GGCAPAENSEPA-----A	2530
QY	3193	KDSOCALENDINELVTGKNNENNEPEKVEYKLTGVLGYFVDETHKNHVDGNTKKEEQ	3252
Db	2531	TDS-----GKLEPQTPVLK-----PEEA	2550
QY	3253	TYRKALFFTPPHVDSTYQAPLESTHRAVOYDKN-----DILKSSISVYIVSALG	3303
Db	2551	VPEEP-----PPPOEKAPAP-----IPQOPPEPTQLDNPVHALVATVSLTASVG	2599
QY	3304	L-----IALHKKKKKSSV-DLRLINIPQEGEMTESEKRRYIPYSGYKDKRTYIM	3358
Db	2600	IGFATFVFLYFKKTKSSVGNLQIILQIKSDIDPITKSPKRYIPYISGKRYRGRRIYTL	2659
QY	3359	EGDT--SGDEDKMWDLSSDITSSSEYEELINDIYVPGSPRYKTLIEVLEPSKR	3415
Db	2660	EGDSTDSGYTDH-----SDTTSSESEYEEMDINDIYVPGSPRYKTLIEVLEPSGN	2713
QY	3416	IPSD-----DPS-----NDTPRTNFIDDEENELKHDFVGOYL--	3449
Db	2714	TTASGNMTTASGNNTTASGNKNTSPDQNDIQNDGIPSKITDNEWNLKDEFTISQYLOSE	2773
QY	3450	PNTEPNNNYTSADIPNTEPNTLYSDNPEKEPFIISIHDRDLYTGEKLSY-----	3499
Db	2774	PNTEP--NMLGYVNDNTHPTTSH-HAVEKEKPFIMSIHDRNLFGSEEVNYDMFSGNPI	2830
QY	3500	NINNSTN-----TNDIPNANANDSYRGIDILINDISLVNLLTY--MMK	3541
Db	2831	NISDSTMSDLSINSHSPINDKNDILSGDILINDALSCHHIDYDEMLK	2880

Query Match 17.2%: Score 3328.5; DB: 5; Length 2527; Best Local Similarity 27.6%: Pred. No. 12e-152; Matches 992; Conservative 421; Mismatches 913; Indels 1241; Gaps 134;

```
QY 31 SHKSARNV-----ERYAKNIRHPSKYAAENHDSLTGDIYTKAEFRGCP---STPNKHNK 82
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 13 SAKDAKEILDMIGEYVEYKEKVKNA---ANDFERIKLGTLSQMPFEAPRKEQOTPCN----- 65
QY 83 YYYPCNDLHKETTNRDYDVNLRHCHGREDNRDEDESECG--NKIRNYRKDAIAC 141
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 66 ----PELXYQWHTNATR--GKNYPCRTGTERSESESGECBDEKIKIDNKGKGG--AC 116
QY 142 APPRRHMDCKNLALANDINTQNIHLLGNVLYTKAGEESLVNHN--PHKGSID---CT 197
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 117 APYRRLHLCVRNLEENISALDKINNPILLADVCLALHBEAALISADHGQYQDINDSQLOT 176
QY 198 ALARSPADIGDIVRGIDMK--PVAHDKYETGLREYFKIHDGME-----DEVKNDYND 250
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 177 MLARSPADIGDIIIRGDLIRGNNGKDKLEENLKITFGNIYKDVTKGKNVALIKRYE-ID 235
QY 251 GSGNYYKLREAMNNVNRKNVWEALICDASYSKSG---YFMOSSNTPPLTSNPKCGKQKV 307
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 236 AIGNYTKLREDMNANRQETIKALTCDA--PBGDVHYFKTCSMGOASHVKRCRLNDP 293
QY 308 PTNLDIYPOLYLRWPEDEGSEFCRKRRIKLKRYKDCSRNDK--ERLCSHNGHDCITTYKK 366
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 294 PTYFDVPOYLTFWFEWMAEDFCTKRKHKLQNLAIKICRGDSGNDRYCDLNGDCTFTANG 353
QY 367 GILHIDNCTDSTCKQVYEWVLGNOOAFKKOKKKYKEI-----QSLISDNKRVN 419
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 354 NKRFSNDEYKCSLSCDHFVPWIDQOKEFEKQKKRYEIKAEKKEKETSITTGDKTIN 413
QY 420 NINSEYKQVEYKLETOYATNDPLNLNBEKYSKG---GLPEKNDITFNSADKGI 475
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 414 NT---YVOEFYQOL--OSSYKDEVEDLOKLEBRICKRPPTYGKEPADVPDTFKGKTER-T 468
QY 476 FYRSEYCOVCPGCGKCGIKYTHKSDMDREVN--NDIYRPMGVKPTNT--VLYSG 530
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 468 FRTYECRCRCMPGCGIK-----KRADGTWEIADMNCVKEKTYEKKNTDIPVLPD 521
QY 531 NBOGDITOKLENFCSNSTYKDKNNQ--KWCYSY-----DENINPKLE--ONT 576
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 522 KQSRLILKKYKFCF-DINGKNSNGQKEMWYKYYKKNKDKDVNDGDSNT--CYLONDQNT 578
QY 577 ELNNNPKILSHNFELMAYTYLLBDITKMDKLTCTINNTTTCID--ECNRNCLCFDRW 635
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 579 BEENDR---SYNSFTFWVTEMILDSIDMKREKELRCTNRNKGKCKNNKCNDDCKFERW 634
QY 636 VQKQEEWNSIJKLTETTKKNKIQOVSYSINNLFGSYFPFKVMDKIDKNOEAKKEL--MENT 693
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 635 IEEKNEMEKIVDYTKYKPGF--DILSNNDYALAKTLILDVEDILLNIEDGKKEVGESEHI 692
QY 694 KKKKNEFSLENNNDYLE-----NAILEDLHLKETATICKDNN--TNEAC---ETS 740
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 693 RKM-----LENEKNOEBAVAAYTVKRTYIDLLEHLEBDADLCLDPLHLEDOCSDDDD 746
QY 741 HNA---TTNCPVVRGCGOPTKNIKELAQYKRSAYEARRGHLK--LKGABEGITKRG 796
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 747 HEEYVYVNPAPAKSGSPSLAN--EAAHOMHKAQKLOLGSRGRRLLKADASGGEYKRG 804
QY 797 GRRKDEKMLCRIMIKHSNRMLGFSNGPCDCKGTGDDIGTFRVVGTGMEVDPEHMRDHE 856
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 805 GOGMTLNDICKITKKNYTN--DIRSNGACGCGK--DGSNBEFKIGTMSVVEDEKKTTSYS 860
QY 857 DVIMPPRRHICSTNLEHLOTDDHPLNGNIVDDLNVNNSFLGDVLSAKTYANKIIRMYKE 916
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 861 DVIYLRPREHEWCTSNLENLDGVSYLENGKAL----HSLGDIYLTPAKMDAAELINRYKS 915
QY 917 KNNLKGEVYDPHQTTCITCAIRYSPADIGDIIIRGDLBERNGDNWKLOGHILEYEGN 976
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 916 QNSIGDP---IDKKHQSISICRAITYSPADIGDIIIRGDIYDNETGMNHLGHLEKFKNI 972
```

QY 977 HKSLSKGNKDYND--PKYLKRENMWEANRAKVMKCDIKYLDKXSHOSTOSS 1033
 Db 973 KEKLPGL-ODKTYGDKRNNPPYKLEEDWEARRHOFWRMCKATP--KDK-----NM 1022
 QY 1034 YGYSDDHPLDDYIPOLKRMATEMAWYCVQKREYDKLKECKECKDKNOGGCTRESG 1093
 Db 1023 KC---NGPIEDYIPOLKRMATEMAWYCVQKREYDKLKECKECKDKNOGGCTRESG 1079
 QY 1094 YGCTCTEACNEYNDIIGLMEQWNTISDKYKELHBOAQMVSNSGIESSTAKNHD-- 1151
 Db 1080 V--CKPAKCKEYKELKINMEBQMRKISDKYKELHBOAQMVSNSGIESSTAKNHD-- 1134
 QY 1152 -RNVIEFLSL---YOONGSKNSKSGTSDESANIGTNTYENGAUYLHDTGNDDCOSQ 1207
 Db 1135 YKQVYVEFLKORTIKSSASKRKRSTDTGTTPTLSPYFAAGYIHOELPMWGCISQT 1194
 QY 1208 EFCDEKSDGDKNEK-YAFRPRKPODHQAGCGSKSPTRQIOTKKKAEKDECKTVND 1266
 Db 1195 HRC-----GDNKKEAFRDKPFDHDKVDCESGER---LOYK-KREGIYLEBPQVYKT 1245
 QY 1267 LKENDGKQVEDCHPK-KNSN-GYPMQCGNINL-VEBPVCMPPRROKLYCHFLANDN 1323
 Db 1246 LKDKNNNSDICEIKRYKGNKGYPMDCRNRKIKIGEGAYMPPRROKLYCHFLANDN 1301
 QY 1324 EIKKLOSQVN--LKEAFISAAETFFSWTYYSKDEGN-----ELDKELKEK 1371
 Db 1302 --KOLDQYDEKLRDAFISAAETFFSWTYYSKDEGN-----ELDKELKEK 1359
 QY 1372 IPAPLRSMEYTFEGDYRDLFGDISKSHGSGSLKEQIDSLFKNGDQKSPNGT---RQ 1428
 Db 1360 IPEEFKROMYTFEGDYRDLFGDISKSHGSGSLKEQIDSLFKNGDQKSPNGT---RQ 1416
 QY 1429 EMTTSHETWEAMCALVYKIGAKKDDTE--NYGINNVKFS-DKSTTLEPAKRPQFLR 1485
 Db 1417 VFMGTGKIDMEGMLALEKITEKQMTFTDKHETFDKVFSGDKSBILEKFAQTQFLR 1476
 QY 1486 WLEWYDDCYTROYKLYKQJEKCKSNDOUKCDECKKCKEDYVYKMKKKEMIPQDKY 1345
 Db 1477 WMTGEBHRCRERKEVEYLIR-----TEC----- 1500
 QY 1546 KDERDKRPRDROHIGYVWTDYTGNTADYLNRRFTASCGDKPGSASVQNHIOLEKQAY 1605
 Db 1501 ----- 1500
 QY 1606 YDADKCGCTKLENDKTYNTISSDKCKGLVEANTGAIKQNKGNPNYNNLKELTEDY 1665
 Db 1501 -----AGC----- 1503
 QY 1666 LFPSSRLRICFHALDGNYPYKENDENGLKRLMEVAATGTYLGOYKKEKKEKIKITS 1725
 Db 1504 -----EVS----- 1507
 QY 1726 DAHRYEYVPPCSAMKYSFYDLNDIILGIDNLEDEKQTEENLKIFPNKGTSGKSGDS 1785
 Db 1508 ----- 1507
 QY 1786 TTGNPOSTARKFTWENKECVMNAMIGYKRGDDGNSGNSASDEDLKCGSVPSDDY 1845
 Db 1508 ----- 1507
 QY 1846 PMGKNDEGTAQFLMPFAWEGEDFCKHKELEKLVGACNDYTCGDNEDKRRKCTDACT 1905
 Db 1508 -----NDSSCHRNGECCCKRECK 1527
 QY 1906 QYKFTISEMPOYEQIKKYGKNDKIYSEHPAKDAEDAREYLDKOLKTI--CENKSG 1962
 Db 1528 KQKMLEKRENTDQKVKFKTDKEG-YNDOPDTIOSTEAYELGKKLKNITFTSGTNG 1586
 QY 1963 DCEYKCKNDVSTORLDGN--SOMNPASLDEPEVEBKCCQYPRGPPRYRRETPSPRVS 2021
 Db 1587 DC--KCMQEKSKQSHSGSITDMNPESLDEPEVEBKCCQYPRGPPRYRRETPSPRVS 1639
 QY 2022 LKSKATSKREAKTAPPTQPKVENVLTETEMRQOTRTBRAAQT RKRKTSTATTESDVT 2081

Db 1640 ----- 1500
 QY 2082 MYKAILSNKPSRSGLEGCPNPTYGQ-----YPMWGTIVGSK--ENENGICMPPRRKLK 2135
 Db 1648 IYDAIILD--PNSGTGYDADCEEKYGNKKTAYPCKKCTINKIEGKEDDYVICPPRROKLY 1705
 QY 2136 INNIQYLVNTEENKRDNDIKAEFIKCAIETOFELMKYIE----- 2176
 Db 1706 IKQKL-----EFGSKNDKELKRAFECAVAETFPAMHRYEKDKKNEEKEEFTSYLVRLGL 1761
 QY 2177 -----NPAENELONGTIPDEFKRIYTYGYDYKDMFGTDISDKKITYTNSVTTI 2229
 Db 1762 GKQOSSPEEBAORDLESKIPDEFKRIYTYGYDYKDMFGTDISDKKITYTNSVTTI 1819
 QY 2230 LNEENKKNODKKDEELRKIFEMENKKEFIEGMIYGTLYLHTEBENEKIRDMY---OY 2285
 Db 1820 F-----PNSVYKPDGQRTREWMNNNAKDYNMICALST--DTKKIKLEGITYAOLTON 1870
 QY 2286 NDMTKLT--PSLEEFVRPOFLRWTEMAEFCNKRREOLKLEAGCK-----EYEC 2335
 Db 1871 NDYKNTYFDNGNTTLEIYATVPQIRSFEMAEFCNKRREOLKLEAGCK-----EYEC 1930
 QY 2336 NGSNDGKTQF-----CAEACVYQNFJAKKTEYERQREKFKDK----- 2375
 Db 1931 SGGHDCTNEELHNDMFADSYCPDCKKACRKYKMLEKVEEYFNQKSKYKREKHLKTS 1990
 QY 2376 --DGKKYKDYPTSERDIEKATCAHEYLNMKLEKLGKDCSQMOPSSQPLKTTQSOSS 2433
 Db 1991 NDNKYKKEFYATSEG--KKSVDSPFLD-KLKE----- 2020
 QY 2434 DANDMPESLDYVPEEFKCECPBELSKSGSMIHTKITEPKIPANCVEKAAVYLSKEAENN 2493
 Db 2021 -----RSHC----- 2025
 QY 2494 MDITLKEFIPISTKEKESKNSMTNNPCDPKPYAPDKYIGRND--GENRENR 2548
 Db 2026 MD-TLEBKT-----DKN-----PLTFSSSTYC-KTCPLVYVQCRNNSDH- 2064
 QY 2549 FKDYEMKCKYKNSKRYOKEKRYCVPPRREHMLRNLEIETIERLKDSNYLLKMYRARN 2608
 Db 2065 -----CIQNSG--NEK--WEHA--LDITIKKNAFYSINVOIMDR----- 2099
 QY 2609 EGIDIINENSEGAMPICDTMYKYSFADLGDIVRGTDMLRIGYLPVPEIKLYYVEY 2668
 Db 2100 -----RG-----OY 2103
 QY 2669 IYKWRKRNKRNKRYNDVOTFRSAMWMDANKDIY--KAMTCAPEDAKLFRGRM 2721
 Db 2104 I-----QEHSENSFESRLKSV--REOKWECGFVKKKDYCV--LKNF-KENI 2147
 QY 2722 DGEFRT--LIDDKCGHKDDPVDYIIPQFRWMTMSE--YYCKALMELEKFKKSCDH 2777
 Db 2148 DTDITITFKVLLN-----WLOPFIGYIYS-----KRIDI 2179
 QY 2778 CKTSDRCKNDYDENKCEQCTRCOEYKNEVYKWSLFDIOSNKKYKLEOPYITKISTYD 2837
 Db 2180 CTKKE-----EHTALEGCKSKCE--CIQK-----LQOKTEWD 2211
 QY 2838 HVOFNQKLTFFSECVSEFSELYHETSKLWAKFENDGSSNIRYAFEEPRKSKYA 2897
 Db 2212 EIKTHRKO-----NRDG-----YETAHKVRYNFE- 2237
 QY 2898 CSCTLPSKNPLJDNCPDQNDKCKELOTFTFCSKNDYDNIMDNNAVLYNNSDNDNGYL 2957
 Db 2238 -----KNAV-----OLKKV-----IDDLKHY- 2253
 QY 2958 IPRRRRLCTRPITAYYRKGDKEILKLLKLTSAFSOGULGQYKKEEELCEPAMKYSY 3017
 Db 2254 ---KIIDDRIRIYVLDACANMKE-TNKDWSIILSO--LKREIKFENQPHETSPNY 2305
 QY 3018 ADYSDIIGTDMDDTSLSEKIKIIFERTSNATEENRKRKWMENNRRQIWHAMLCGYKIATSK 3077

```

Db 2306 CDISPHILTEPSOTDTEINPYDET----- 2332
Qy 3078 VTLDEGWCQLPDEINOFLEWLEMAKQACEKKEHVSDSLKTCPRSNEDNEFEASELLR 3137
Db 2333 -----PEDDTS----- 2347
Qy 3138 QPGQONDRKATISANILIKNTMENLNTKYKOLKQDSSGNIDNKPSEENVOSYIKSKDSQC 3197
Db 2348 -----PEDDTS----- 2376
Qy 3198 ALLEINDINEIYTGKNNENNEFEVLEKLYPGLFVEDETHKNVLDGNKKEEQIYRKP 3257
Db 2377 AAP-----PVAPSGSEEEAPKEVPEKRP-----KEVPKPGPK 2409
Qy 3258 ALIYFTPHVDSFYAPLFSTHRVAQVDPKNDILKSISV-----VIYSALGLIALHFMKK 3312
Db 2410 A-----PKRRPREVTH-----SILPEWVISAPLPSVGLIAPALISYFLAK 2451
Qy 3313 KFKSSVDLRLINPOGEVMPLEESKNRYIPYRSGPYKRTYTYMGDTSGEDKYMMD 3372
Db 2452 KTKSTIDLLAVINPKGDYIPIPMKSKNRYIPYASDRYKGTIYMEDGS-DDYTYIGD 2510
Qy 3373 LSSSDIT-SESEXEEL 3388
Db 2511 ISSSDITSESEXEEL 2527

RESULT 9
ID 025733 PRELIMINARY; PRT; 2924 AA.
AC 025733;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE PFEMP1 variant 1 of strain MC.
GN MCVAR-1 PFEMP1.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MALAYAN CAMP;
RX MEDLINE=95330812; PubMed=7541722;
RA Baruch D.I., Pasloske B.L., Singh H.B., Bi X., Ma X.C., Feldman M.,
RA Taraschl T.F., Howard R.J.;
RT Cloning the P. falciparum gene encoding PFEMP1, a malarial variant
RT antigen and adherence receptor on the surface of parasitized human
RT erythrocytes.
Cell 82:77-87(1995).
EMBL: U27338; AAB60251.1;
InterPro: IPR000345; CYLC_heme_bind.
DR Pfam: PF03011; PFEMP. 2.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
FT VARIANT 104 104 S->T.
FT VARIANT 105 105 E->K.
FT VARIANT 139 139 D->G.
FT VARIANT 362 362 K->E.
FT VARIANT 503 503 K->N.
FT VARIANT 581 581 S->T.
FT VARIANT 593 593 M->V.
FT VARIANT 617 617 I->K.
FT VARIANT 631 631 V->G.
FT VARIANT 658 658 D->E.
FT VARIANT 722 722 K->N.
SQ SEQUENCE 2924 AA; 335860 MW; 7077553BDAC82B26 CRC64;

```

```

Query Match 16.18; Score 3130; DB 5; Length 2924;
Best Local Similarity 26.78; Pred. No. 6,1e-143;
Matches 1005; Conservative 445; Mismatches 968; Indels 1350; Gaps 151;

```

```

Qy 33 KSARVLEERYAKNT-----RHPSKYAKENHDSIAKGLITKAEFGFSTPYNKNKNTYTPY 86
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

```

```

Db 13 KQAKHALDRIGEVEYKEVNAEKYK-----ALKGNLOEANGIGLASS-----PN 60
Qy 87 PCMLDKHEHTN-LRYDVLNRHPCHEQONRFEDESECG-NKIRNYRKKDAI-ACAP 143
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 PCKLVEDYYNNRLK-----RRKYPCANRQTVRFSDYGGGCTNRILKDSNNDSIGACAP 116
Qy 144 PRRRHGDKNLELNDINONIHDLGNVLYAKIGESIVYNN-PHKGCT-----SDACT 197
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 117 YRHLHLCIDNLEKMGKTSSTK-HDLLDVCMAKKEGSIKHYHRTKHELINDTKTSQLOCT 175
Qy 198 ALARSPADIGDIYRGIDMF-----KPNVHDKVEYTGRLREVEFKINDGMEDEKKNYNDG 251
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 176 ILARSPADIGDIYRGIDLVGYDDKKEKDERKLENNLLEIFKIHENLGTQAKHYKRD 235
Qy 252 SGNYKYLREANNNVNNKWEALITCDASYKSGYFMQESNPLFSNP-KCGHKQK----- 306
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 236 EENYVQLREDWMTANSTYWKAITCHAGSDSYFFRTCCSGEMTIDKCRKDEEGNETN 295
Qy 307 -VPTNIDVYPOYLRFMEDEWEEFCRRRNKIKLVKVDSCRNK--RLYCSHNGHCTTTI 363
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 296 EYPTFEDVYPOYLRFMEDEWEEFCRRRNKIKLVKVDSCRNK--YCDLNGYNCETA 354
Qy 364 WKGLIHLNCKTDCSTYCKVFEVWLGNOEAFKKQKREKEI0---SYLSNDNKFVN 419
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 355 RGAELIFVKGDDCHKCSVACDRFVKWIDNQRKDFQKKKYDEIKTHTTITTINGKIN 414
Qy 420 NINSEYKQFYKLEKETQYATNDTFLNLNBSGYK--GGLPEKDTFTNSADKGIY 477
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 415 NL---YGHFKILKK-YYPVDSLSQRLNDEALICKPPNGNKAASYDNNVNTFS 470
Qy 478 RSEYQVCPDGVKCD-----GIKYHKS-DNDRREVNNDEKPPMGVYPTNITLYSGNE 532
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 471 HTTYEACPPMGAGQKEKNGGAKKCKAKKERRIFKE-----NSIDIKILTPBK 523
Qy 533 QGDITQKLENCNSSTYKKNKNOKECYDENINRCKLEONTLEIND----- 581
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 534 RSKLEKLEKTCCKQKQKIK-----NDIMKCHYDDNG-----TDQDSDSDVLGDMGLTK 575
Qy 582 NPKIISFNFELWYTYLLRTIKRNDKLTCTIN-NTTHCIDCNCNCLCFDRAVYQKE 640
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 576 EDKIMSYNAFFMMVYHMLIDSIKRDEHGRCIKDKGKTCICGNKKJCICFQWVYQK 635
Qy 641 EEWMSIKKLPFKKKNIDQOOSYSNINLFEYGFVKMDLKDKAKEKLEMINIKKKEE 700
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 636 TEWKIKDHFQKQDIPDM----- 655
Qy 701 SNLENNRDYLENAI-ELLIDHLEKETATICKDNNTNACETSHNATNPPCVKPRGCTOPT 758
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 656 ---THDDFLQTLMKDLLEIIDDY-----GDANE----- 683
Qy 759 KNIELAQYFERSAYEARNRGLKLGKAHEGYKRGGRKDFKDLICRIMIKHSNRL 818
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 684 --IKRI-----EALLDQA--GVGGIDFALAGIYTGFAE--KOTITDKLLQHECK-- 729
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 819 GFSNGPDDGKGIGDGTQTRVVGITGEVDEPHMRKDEHYIMPPRRRHICTSLMHLQTD 878
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 730 -----EADKCLKHTIDTQCPQ-----E 747
Qy 879 DHPLENGIVDLVNNSEFLGDLVLSAKYEANKITIRYKKNLKGPEYVDPKHQTITCA 938
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 748 DRSY-----ARSESATV-----PSPPADPAKTEEV--- 772
Qy 939 IRYSPADIGITINGRDLMEWNGMVKLQHLTFVGNHRSILGKGNDRYNDAPRYLKL 998
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 773 ----- 772
Qy 999 RENMMWEANRAKWEAMKCDIKYIKDSGHSOSTSYCGYSDHTPLDDYIPQKLRMTWEA 1058
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 773 -----DNNS-----SDEE--DDFEE-- 787
Qy 1059 EWYCKQKKEYDLKKECKCKDKDNGGCTK---ESGTGCTKCTACAGNEYDIIGIMKE 1115
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 788 -----EEEBEDGGEAEAEVGEKTEDESATRAVAPSPGCTQ----- 824

```

QY 1116 QWNIISDKYKELHEQAOVSNSGIEASSTAKNHIDRNVIETLESELYOONGKSNKSGTS 1175
 Db 825 -----DGVPAQOEEDVAKCSI-----VDKALKGRLDADCTLYKGT--APTS 865
 QY 1176 DESAVIGTNTTYNNGAYLHDTGNFDCCOSQNEFCDEKSDGKNEXYAFADKQDHDGAC 1235
 Db 866 WKCIIPSGNNTTEST-----TKP-----GAA 886
 QY 1236 GCSGSKPTRYOJTKKKAERKDECKTVDILKENDGKKQVEDCHPKKNSNCPMDKQ 1295
 Db 887 GTPSG-----KDT-----G 895
 QY 1296 NINLVEDPRVCMPPRRKOLCVHFL-----ANDN----- 1323
 Db 896 SI-----CVPRRRKLTVGKLDHAGGETTEAKSOFTSGGOKTPSGNESSPSEKLPQ 947
 QY 1324 -----EIKKLOSOYNLEAFIKSAAEFTFSWYIKSK-----DGEK 1360
 948 GPTPETTKETPESLLHAFVSPRLRFLFMHFKEDQMKAHAGAGTQOITIGLDGGG 1007
 1361 NEL-DKELKGRKIPPAFLRSMFTYFGDYRDLFG-TDI-----SKGHGSKLKOIDS--- 1412
 Db 1008 EEPFDLTKGHIIPROFLRGMFTLGRDILVGNVDIVHTSGNKEDQIMALQKIE 1067
 QY 1413 --LEKNGDOKSPNGKT-----ROEMWTEHSHETIEMALCALV----- 1447
 Db 1068 OILPTSGSSPSPPRYVOTQOHSVENPRKTMMENNGKIMEGMVALVNTDTPSGTAPTOI 1127
 QY 1448 -KIGAK-KODFTEN-----YGYNNKFS-----KTTLEEFKAPPOFLR 1485
 Db 1128 QERTLROBNSKNKPIPOKRYDOVKLDITSDAKTGSPPVSGEKTPTPLDFTSRPYFR 1187
 QY 1486 WLEWTDYCYQTKYKLDKVOEKCK-----SNDOLKC-----DTECNK-----CEDY 1528
 Db 1188 YLEMBETPOKEKREKREKREKREKREKREKREKREKREKREKREKREKREKREKREK 1247
 QY 1529 VTKMKKKKKEM--IPQKYYDEBDRKRRDROHIGVMTDYGYNANDYLNRRKFTASGDK 1586
 Db 1248 HIOCRKRYKRWIDIKFEYHKOER-----KYQG-----EY-----DK 1278
 QY 1587 PGSASVQRNIOLEKQAYYDADKHGCGTFTIENDKTYTNISSKDKCKGLVKEANTGAIR 1646
 Db 1279 -----LTKDKSSGGDNCC-----CKDIEKHSAAVF- 1304
 QY 1647 WOKGPNYNNKLELTDVLFPSRLRILCFHALDGNVTDPEVKDENGILRRKLEAVATEG 1706
 Db 1305 -----LKEIK-----HCKNGOTSEKNGQEDOL- 1327
 1707 YNLGOYKKEKKEKKEK-KTSDAHKYSTYEVPCSAKYSFYDLNDIILGIDNLEDEKQTE 1765
 Db 1328 -----NKLDPDKIPOPSPSTYKACP-----YGV- 1353
 QY 1766 ENLKTIENKNGTSVGKSDS--TTNGNGSARKEFMNENECEYWNAMIC-----CYKR 1016
 Db 1354 -----NCGNKRGRGKTNGCTTNE-----PENENKGAASTISILINDSTN 1397
 QY 1817 GRDDGNSGNSARSDDEDLKKGCVSPDDPYM--GKNDEGTAYOFLWFAEMGEDFECHKR 1874
 Db 1398 GAINGTGT--IDETLKEC-----SDKVAFFGLRQOBTQCK-----KYGVNC-- 1440
 QY 1875 EKELEKLVGACNDYTCGDNEDKRRKCTDACTOYKKFISEWKPOYEQIKRYGENKDKIYS 1934
 Db 1441 -----NLTNRYND-TYEDKO-----IVNEFEFORLRYF--VHDYINILKHKI- 1479
 QY 1935 EHPVADADEAREYLDKOLKIKKICENKSGDEYECMDVSOURLDNGSOMPSLDEPRK 1994
 Db 1480 -----DPCIKKEKODKT--EHKCI-----NGCNI----- 1501
 QY 1995 EVEGKCNCOVPRGPRRYRRETPSPRVLISKATASKKEAKTAPPTQPKRVENULTTEMA 2054
 Db 1502 -----KCCSC-----VR----- 1507

QY 2055 QTRFRRAAOQTRKRTSTATTESDVGTMAKAILSNKPNDSRGIEGNCPKTYGOYPRKGC 2114
 Db 1508 -----KWLEI 1512
 QY 2115 VGSKENENGICMPRRKRIKILCINNIOYLNTEENKRDNDKEAFICAIETOPMLKTY 2174
 Db 1513 KG--NEWG-----NIK--KHYNINSNDKERT----- 1534
 QY 2175 IENPAENELNGTIPDEFKRIWYTYGYKDM-FRGTDISNDKKIITVNSVTILINEN 2233
 Db 1535 -----IAYNKSYVDGCLFDTDKKAKQVY----- 1560
 QY 2234 NKKKODKKDEDLKIFEMENKKEFTWEGMITYGLYHLTDE-NEKEKIRONYOYNDWKLT 2292
 Db 1561 -----EDEKERRKTIW-----GCTGH--DECSBEKEENK--NFTNL 1594
 QY 2293 PSLEEVKRRQFLRWFTMAEFCNKRKBDLLLEAGCKREYECNGSNDGTOCAEACVT 2352
 Db 1595 SELQDKITSCQ-----NNH-----NPKGT-----AC-- 1616
 QY 2353 YONFIKWKTEYERQREKFKDKDKKDYKDYPTERDIEKATCAHEYLNMKLKELCGNKD 2412
 Db 1617 ----- 1616
 QY 2413 CSCMQRPSSOLPPTQOQSANDMPESLDYVPEEFNKECEBELSKGSMHTKTIPT 2472
 Db 1617 -----DPPSPTEBETPLDDDPDPLDDQHTBQKFCP----- 1653
 QY 2473 KIPMNCYKAAAYLISKAENNDITLKEK-----PIPIESIK-EKESK 2514
 Db 1654 PPRPTCYEKLAKELEYAEKINNELKNGKDNFGKCNVKKNGAVIGESCKFEQTYE 1713
 QY 2515 NSWTNNPCOPKRPYAPDKYIGRRNPENRENEFRKDYEMKCYKNSKFOE--KRRVCP 2573
 Db 1714 NSVNNIN-----NCKDONERFKIGQW--NKYIGTIRKDCIP 1752
 QY 2574 PRREHMLRLDEKIERLKDOSNYLKMVRRTABNEGIDILKFNSENSEGCAMPIDCTMK 2633
 Db 1753 PRREHMLDLDSMIGRTTIDSSALLKIOEAKSEBDIIRKLLEQNSCDEHRICDAMK 1812
 QY 2634 YSFADGDIYRGDMRLRIGYLPVPEIKLYKVEYIYGGKRN--KN--KARNYNOVQTF 2689
 Db 1813 YSFADGDIIRGRDLNKNKSKQGLQRLYATLNTYKLNQDNKYEKRPYLOD-- 1869
 QY 2690 RSAMWDANRRDIWKAMTCAKAPEDAKLEFRKGRMD-----GFERITLLQDCKGHKDDPEVD 2744
 Db 1870 RSDMWDANRRKHINNAMCNAPDAKFLKKNPNDTSSSSSKGIMTHTSNGYKPEPDYD 1929
 QY 2745 YIPQFRWMTWESYICAKALMELEKFKSCDHC-KTSRCKNDYDENKCEOCKTRCOEY 2803
 Db 1930 YIPQFRWMTWESYICAKALMELEKFKSCDHC-KTSRCKNDYDENKCEOCKTRCOEY 1989
 QY 2804 KNEFLKWSLFDIOSNRYKLEYDQPIY-----TKISTYDHQNVOKYKTKRSKC-SVESF 2858
 Db 1990 KKLHNMKLFCD-----KIKTIYNE-IYNNKOSKINSNEYKFKLEKL--KDKCKELNSS 2041
 QY 2859 SEYLHEKSKLANKFNENDSSNIRTYAFETEPKSYKEACSTLPSKNPLNDPTDQ--N 2916
 Db 2042 DKCIDENATHCYTKKFS-NSENKNNHNNVAFKNNPREYKAKACKOAP--DPLDNCPRKSDATY 2098
 QY 2917 KDGCKELQTFPSCSKNDYDNNLDNMAVLYLNSDDKGVLLPPIRRRHLCRPTIAYNRY 2976
 Db 2099 EKACNTLPLTKIESKTFNNDDSDMTSFYQTSRDTGVLPVPRRROKICLNITTY- 2157
 QY 2977 KGDK-EIKKRLKLTSAFOSOLIGOKKSEELCFEMAKSYADYSIITCTMDMTSLS 3035
 Db 2158 SIEKIDFKALMTSAVNEKSLCELTCKDRDYLOAMKSYFYDYGIVAGTDLISTAPL 2217
 QY 3036 EIKIKIIP-----ETSNEAENRKTWNNENRROIWAHMLGCIYIANSKYVLDBGWCOLPK 3089
 Db 2218 DKLTKRLNVLKGDGTRNEIEDRGKMTWENTRYWAMHMLGCIYAAAGKI--EERDCSLP- 2274
 QY 3090 DEETNOFLRWLITWAKOACEKKHVSLSLTKCPKRSNEDNEFASELLROPGCO--NDIR 3146

```

Db 2275 DONTHQLRMREWESEHPCARQKLFNEVREC--ASAOCIIIEGTIDPVCCEACTQYR 2332
QY 3147 KYISLNLILIKNTMENLIRKRYKOLKODSSGNIDNKPSE-ENWQSYIKSGDSCALFLINDIN 3205
Db 2333 DYIRKIQ-----EYRLNLYQYINFNENKKALEYTAPEYFNDK---CNDKNCJLS 2379
QY 3206 EYVGTGNNNNENPEKVKLILYVEDETHKNNHLDGNIKEEQTVPRKALYFTPH 3265
Db 2380 KYI-----DIEKKMKMNYDSF---DDNDLKNKCICQIKPKRP---PKVKKPEEH 2424
QY 3266 VDSFYQALPSTHVAAYD-----KNDILKSSIVYISALGIALHMKKKFSSV- 3318
Db 2425 TPSSQDTPLPPPKPDDLPPPAEPPNBDILETIPFGIALALGSLAFLELKKRKYSSVG 2484
QY 3319 DLRLILINPOGEGMPLLESKNRNIPIYRSQPKTYIEMGDT---SGDEDYIMDLSS 3375
Db 2485 NLFQILHFKPSQYDIPKLSPNRYIPTSGKYGKRITYLEGSDSGSDGTIDH----- 2538
QY 3376 SDIT-SESEYEELDINDIYVPSPKYKILIEVLEPSKADIPSD--DPSS-----ND 3425
2539 SDITSESEYEELDINDIYVPSPKYKILIEVLEPSGNNNTASGKNTPSDTQNDIQND 2598
QY 3426 TPRTNFEIDEMNELKHDEFSQYL---PNTPENNNYKSADIPTNTEPNTLISDNPEKPE 3482
Db 2599 GIPSSKITDMEWMLTKDEFISNMLQNEPNTPE--NMLGYVDDNNTIHTT--SRHNVKEKP 2655
QY 3483 IISIHRODLTYGKEISININMGSTNTNNDIPMARNDSTYRGIDILINDSL 3530
Db 2656 IMSIHRODLTYGGEYSYVNM---VANNIPISARNGNYSIDILINDSL 2700

RESULT 10
QY 0904A2 PRELIMINARY: PRT: 1685 AA.
AC 0904A2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-MAY-2002 (TREMblrel. 20, Last annotation update)
DE Variant surface protein PfEMP1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_Taxid=5833;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE-20144115; PubMed-10677532;
Smith J.D., Grais A.G., Kriek N., Hudson-Taylor D., Kyes S., Fagen T.,
Pinches R., Barruch D.I., Newbold C.I., Miller L.H.;
"Identification of a Plasmodium falciparum intercellular adhesion
molecule-1 binding domain: A parasite adhesion trait implicated in
cerebral malaria.";
PL Proc. Natl. Acad. Sci. U.S.A. 97:1766-1771(2000).
DR EML: AF193424; AAF18980.1.-.
DR InterPro: IPR00454; Eub_Atpase_Csub.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP. 1.
DR PROSITE: PS00605; Atpase_C_1.
FT NON TER 1685
SQ SEQUENCE 1685 AA; 193671 MW; D1PD42666B055IE CRC64;

Query Match 15.1%; Score 2932; DB 5; Length 1685;
Best Local Similarity 38.6%; Pred. No. 1.2e-133;
Matches 650; Conservative 265; Mismatches 619; Indels 152; Gaps 51;

QY 33 KSARVILERYAKNIRHPKRYKAKEHVS-LKGDLYTAERGGSPSPVNNKNNHYTPPCNLD 91
Db 8 RSARILIEFALLIQQAHNAAEQYHKVYLQGDLYKVPFSKGSVYTD-----PCKFD 60
QY 92 KHEHTNLRYDDVNLHPCHGREGORNFDEDESEC-GNIRNYKKKNAL-ACAPPRRHH 149
Db 61 HTRDINVTIGQ-GREYFCFGQVRFSEDRSQCMSNRIRG--SENDVYGACAEYRLHIV 117

QY 150 CDKNLEALNDINTQNIHOLLGNVLYTAKEGSYNNHP--HKGSIDACTALARSFADIG 207
Db 118 CDKNLEQIDPAKITTTNHNLLVLDVLAKEHGESIIDNYPSDHNNKRGICTALARSFAD 177
QY 208 DIYRGIDMEKPNVH--DKYETGLREYFKIHDMGEDEVANDNPPGSSGYTLRLRANVY 265
Db 178 DIIRGIDLVRGNKQOEKIDQDLKTKFKELHNNLSEANNTN-DTKKFFELRDMWNA 236
QY 266 NRNKWEAITCDASYKSGYFMQSESNTPLPSNPKGCKHOKGKVPNTLDVYQYLRWFEDWG 325
Db 237 NRDLIKALICTAPQAYFRNACGGEHTMTYGQRCVNGEPTNFDYIPQYLRFEEMA 296
QY 326 EECGRNRNKLKLVKVDSCNDKMERLYCSHNGDCTTTTKKSTILDLNKCIDCSYCKYF 385
Db 297 EDCRIRKQDLQNVNKSCHGEDNNKYSNGYDCQTLNKIGHRRFGNCJCLCEPCAPY 356
QY 386 EYWLGNQOBAFKQKQEKYEKEIOSYLSNDNRVFNINSEYKQFYKELKETQYATNDFL 445
Db 357 VAWGNKKEDEPKQKCKKENEIYKNNKNOTGSSISVNDMYIDHFTSLAKE-KYSSIHFT 415
QY 446 NLLNGKYKGGLPBEKIDITFTNSADKGIYRSYCOVCDQCV--KCDGIRKYHNSDN 503
Db 416 NLLNBAKCK-MLNEDNESKIDFNDITFTSGSOYCKPCECGEYKKSOG-SFRVRDKE 473
QY 504 DDERVNNEDYKPPMGVCKPNTNITLYSGNEQGDITQKLENFCNSSSTNYKDKN-NQKMECY 562
Db 474 EDACHKHNPYTPNNQVCKPTEINVLSSGEGLEDIEKLEPKCTSDNNKDFSLYEBKCY 533
QY 563 KDEINRCKLEONTEINNDNPKI-ISFNFEIWLTYLLRDTIKNDKLTCTINNNTTHC 621
Db 534 EHDGNEACILKKEKSKRAEIQKSFDFFTYVAHMLKQSEDMKTLSTCLNKQKQDLC 593
QY 622 IDECRNLCDFDRWYKQKEEMNSIKKLFTKKNIQOSYSININNLPEGYFFKYMADLK 681
Db 594 INNCNRKCTEYKTEKEEEMEKIKVHEYKQKPEGHANYILGAYLEDFIKGITAAYG 653
QY 682 DEAKKELMENIRKKNFSLNENRDYLENAIELLDHLEFATICDONTNNA-C--- 737
Db 654 EKEVQYIAELLRNKTKGPDQATQ---KTIIDELDHLEBAHECHRENNPEKNSKE 710
QY 738 -----EISHNATNPCKYPRGQIOPYKNIKELAOYFRSAVEARN-----R 779
Db 711 PHDDIDDEDETHY---NPGCKTDG---TVRAQIKAKKFORDAKQMKNTRNDGTGRK 763
QY 780 GLH-KLKGAHAGIYKRGRRKDPK-DNLCHIRIMIKHSRNLGFGNGCGDGTGDTG-- 835
Db 764 GAHNSIVGDISKAFPKNGGSDLDKDKIDINTSHSDSGNGGPGCTGK---DGNQGG 820
QY 836 TRFYVGTWEVDPPEHMKRKHEDVIMPPRRHICTSNLEHLQTDHPRLNGNIVDLVNSF 895
Db 821 DRMKIGTPWSKVGD-KTYTSDVYLPPRQHMCTSNLEFLETQKTPLDGKFGVDKINHSF 879
QY 896 LGDVLISAKYFANKIIRMYEKNLNLKGPKEVTDPKHDTTCIRARFYAFADIGIIRKOL 955
Db 880 LGDVLIVANPEAKIKELRYKNNND---RKDLNDANDLEYICRAKKSFEADIGIIRDTM 936
QY 956 WERNGDWVLOGLHETVEFNGIHSLKSGKNDKYNDAPKYLKIRENWEANRAKVAEMAK 1015
Db 937 WDKDEGSKKMDVILKILFKGIKQELPKELQKTKNPNPGKITQLRKDWENENRQVWRAMK 996
QY 1016 CDIKYLLDKSGHSGTSSYSGSDHTPLDYIPQKLRMTWEMAWEYCKVQKKEKLEK 1075
Db 997 CAL-----DQGSIEKC---NGIPLDYIPQRLRMWTEMAWEPCMKQKEAYNELKGR 1044
QY 1076 CKECKDKNDGQCTKESGCTCTKTEACNEYNIIIGMKRONNLTIDKXELHDAQMSY 1135
Db 1045 CSQCKTMD--KKCTNS-DCDNTCTEACTAYNNKINTWKQOMALISDQIOFLYLAKTA 1101
QY 1136 SNSGIEASTAKNHNDRNVTIELSELYOONGK-SNKSQTSDESAYI-----GTWTTYE 1188
Db 1102 ANGGPHASSGDVGKDKPVVNFLELYKQNGKIRISPSDHPBRVYKRAPSGNSNTVYS 1161

```

QY 1189 NVGAYLHDTGNFDDCOSQNEFCDEKSDGD-NKTYAFROKPOHODGACGSKSPTRVQ 1247
 DB 1162 TAAGYIHQEAHIDCKNQNVCERKGGNNENKTYAFHEPYPDHKACACDGRNPDVY- 1220
 QY 1248 IKTKKAEKEDCTEYNDILKENDGKQYEDCHPPKNSNGYPMOCGNINLYEDPVCM 1307
 DB 1221 -----LEDECKAVOKILISFOIEKNNIHC--KTEDA--KWGCENTKLEDEGVC 1267
 QY 1308 PPPRQKLCYHFLANDNEIKKLOSOVLKEAFIKSAAEFFSWYYSKSDGEGNELDEL 1367
 DB 1268 PPPRQKLCYHFLANDNEIKKLOSOVLKEAFIKSAAEFFSWYYSKSDGEGNELDEL 1321
 QY 1368 KEGKIPAPLRSMTFPGYRDLFTGDISK--GHGEGSLKQOISLFPNGQSKNGK 1425
 DB 1322 HRDMPPEFRSMFTYFGDRDICTDISEKTIADHDYTAKKRTAVFQKISKTNGK 1381
 QY 1426 ---TROEMTEHSEIWEALCAL-VKIGAKKD-----FTENYGNVNFSDKSTLEE 1476
 DB 1382 KYLEREGMKYELSTKMGICALSYNETKKNDEGVRYLAKYIKN---NDIKETLEE 1438
 QY 1477 FAKRPOFLMFLWYDGYCYTRQKYLKVOEKKS-----NDOLKCDTE-----CNK 1524
 DB 1439 FASRPPELRYVTEGDEYFNKREKELVSLKCKDCSLRNNGTSNKTCDNENCGACKTQ 1498
 QY 1525 CEDYVYVMMK-KKEMIPQKYYKDERDKRFDQOHIGVMTDTYGNATYLRKFTASC 1583
 DB 1499 CEKYYKMMERKHYSSQKKFQLYKNSATYNN---GLAVKE---ANSEYKNDPEVTEA 1552
 QY 1584 GDRPGSASVVO---RNQILLEKQAYYDADKHCCTKFIENDD---KYTNISSKDKCKG 1635
 DB 1553 NSAKHARDYKLPQLENNICNGMYKNCQDYTCANTSSSTNSBPASLDITPSEYKDKCNC 1612
 QY 1636 LYKCAN 1641
 DB 1613 VPDECS 1618

RESULT 11
 ID 087560 PRELIMINARY; PRT: 2658 AA.
 AC 087560;
 DT 01-JUN-2002 (TREMELREL. 21, Created)
 DT 01-JUN-2002 (TREMELREL. 21, Last sequence update)
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)
 DE Erythrocyte membrane protein-1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 NC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5833;
 RP SEQUENCE FROM N.A.
 RC STRAIN-JDP8;
 RA Chantapadhyay R., Pillai C.R., Chluis C.;
 RT Identification of a domain responsible for binding to intercellular
 adhesion molecule-1 from a Plasmodium falciparum field isolate.
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBD databases.
 DR EMBL: AY028643; AAK49742.1;
 FT NON_TER 2658
 FT SEQUENCE 2658 AA; 301326 MW; AEE33BC9167E7B7 CRC64;

Query Match 14.8%; Score 2869; DB 5; Length 2658;
 Best Local Similarity 27.6%; Pred. No. 2,3e-130;
 Matches 869; Conservative 366; Mismatches 906; Indels 1012; Gaps 122;

QY 31 SHSASRNVLERYAKNIHPSRYAKEHVSLSKD-----LTAKFRCGSPYVKNHYY 84
 DB 11 SGTAEFLDRIGEV-----YKAHDAKSGMGLKGLTSLQALFEKAPKQKOTSEB--- 62
 QY 85 PLYCNDLKEHTNLR--YDVNLRHPCHGREQNRFEDESECG-NKIRYKRRKNDALAC 141
 DB 63 --PCDLNHEHTVTVSGYDREN--PCKDRPEVFSYTEGAECDSKIRGSNSNKGQ-AC 116
 QY 142 APRRRRRMCKNLEAL-NDINTQNHDLGNLVLTAKYBESEST-----VNNHPRKGT 192

DB 117 APRRLHLCDPHLEHKKHDTTR--HNLLADVEAAKFAESESLEKRGQYOLNN----- 168
 QY 193 SDA-----CTALASFADIGIVIGIDMFKNVDK-----VEGLREVFKKIH-----G 238
 DB 169 SDVININCTELARSPADIGIVIGKDLFLQNTYESAORILLENKKLEIKPIKIHSDVTKG 228
 QY 239 MEDEVKNDYPDGSGNYKYLREAMVMVNRKWEATITGASYSGYFMOSESNTPLSP 298
 DB 229 TNGELKKRYEKDQ--NEYQLREDMMVTANRETYVATITGAGTSDKRYRNTCSNKGATSG 287
 QY 299 KCGHKQGVPTNLDYVQYLRWFDEWEECRKRNILKRVKQSCR--NDEKRLYCHNG 356
 DB 288 KCRQNDQVPTFYDYVQYLRWFDEWEECRKRNILKRVKQSCR--NDEKRLYCHNG 346
 QY 357 HCDTITIKKGIHLDKCTCSTKCVFVWLGNOEAKKQKQKKEKELIOSYLSNDK 416
 DB 347 YDEKTAAGKHDFEDDYCDQSCAPFDYLDNQLKLETKORNKTEKEMQY-TNGTK 405
 QY 417 FVNNINSE-----YKQYELKETQYATNTFLNLMEGKYCK-----G 456
 DB 406 ---NSKRKKPGAGSKSYNDYGERFYNIL-EGYNNVDFLIDLNETCKKNNEIEG 459
 QY 457 GLRGEKDITFTNADD--KGIFRSEYCOVCPDCCYKCGITRY-TKSDNDEKRVANNED 512
 DB 460 GQINFKNVNSGKNSDDDSKRTTCRTTYCOACFWCGAEEDNSGKWKAKANDADCCGRKD 519
 QY 513 YKPPMGVPTNITVYLSGNQGDITOKLEFNCNST-----NYK----- 551
 DB 520 YDPE---KTTIELIGDTRKSDMYQKFKCTSATCATAPDTAATGEGNGKAGASGNG 576
 QY 552 DKNQKMECYKQDEN-----INRCKLBQNTIENDNPKIISFHFELWYTYLLRDT 603
 DB 577 DNITEWTQYKYYKKNQKDYVKKDINFVLQDQK-HTKEQKVTSYNVEFWKMYDMLHS 635
 QY 604 IKMNDKLTCTCINMT-THCHID--EONNCLCPDRWVQKKEEWSNYSIKLE-TKKNNIQS 659
 DB 636 IKMNEELRSCINNAKSONCKNNKCNDSGCFEWEIQKKEKDKDKHKTQDFGIPGG 695
 QY 660 YYSNINNLFEYGFYVADKLDKDEAKKKELEMINIKRRKNFNSLENNRDYLENA----- 713
 DB 696 PLOQFD-----YDFVLKFTYDK-----KELLQNKOTHAQAKDGRLEKMLEQAGVDSG 745
 QY 714 -----IELLDLHLEKATYICKDNNTNACETSHNATTPCVKPRGCTQPT 758
 DB 746 SGGGNGAKGKHNTKIDKFLQEEQFAETCK--QTONCPKPKTKYANPCY---GNNTYD 800
 QY 759 KNKKEIAQYKRSAYEARN-RGLHRLKGAHGIYKRGGRDRDKNCLRIMKHSNRN 817
 DB 801 ALAEKVAQILQGEAQTLINHRSSSLKGDPEQGHYNGNKNKSVLKD-VCKITNQYSNAG 859
 QY 818 LGFSNGPDCGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDTGDT 877
 DB 860 KNSNDPCNCKN-----ENRPNIGEKKNGCEYKASTHS-YMPREHCTSLNLEHLK 913
 QY 878 DDEPLNGLNIVDVLNNSFLGDVLLSAKYEANKIIRMYKERNLKGPEVTDPKHQTICR 937
 DB 914 DKGRFEOVPDNKATLSLDLVAARKKAEKIKELYEKKNKQSGKNAKGNLNEKTYVCR 973
 QY 938 AIRYSPADIDDIIRGRDLMERNNGDMVLOGLHLEVPENIHSLSKGNND--KYNDAKPY 995
 DB 974 AIRSPADIDDIIRGRDLMERNNGDMVLOGLHLEVPENIHSLSKGNND--KYNDAKPY 1027
 QY 996 IKTRENNWENRANKVWEMAKCDIYLYLCKDGHOSQSSYSGYSDHTPLDYIYPOKLEWMT 1055
 DB 1028 LELRKDMWENRANKVWEMAKCDIYLYLCKDGHOSQSSYSGYSDHTPLDYIYPOKLEWMT 1075
 QY 1056 EWAWEYCKVQKKEKCKECKECKECKECKECKECKECKECKECKECKECKECKECKE 1114
 DB 1076 EWAWEYCKVQKKEKCKECKECKECKECKECKECKECKECKECKECKECKECKE 1131
 QY 1115 EOMNIIISDKKELHEQAOQSVNSGIGASSTANHNDRNIEPLSELXYOQ-----NGK 1168


```

Db 1132 QQDMKMLKTYLLYANKTTSTNAGRTVLGDSAPDY-QQMLDEFKLOKEIKNSALNRTK 1190
QY 1169 SNKSGTDESASVIGTNTTYENVGAYLHDGTGNFDDCOSONEFCEKSPDKNENKAYAFRDKP 1228
Db 1191 RSJDTGNNDPIF---TSPYSSAGYIHQELPHYQCOVQKFCYNTNG---NKEKTYARHPD 1244
QY 1229 QDDHAGCGSGSGSKPT-----RQYITK-----KKAERD 1258
Db 1245 HDTDEACACRPSTPESPANLPLADOGKYAEAEAEEDDEEDDDDDDEDEDEDEED 1304
QY 1259 TECKTVNDILKENDGKQVEDCHPK-KNSN-GYPMOCGNINLVEDP-RVCMPPRROKIC 1315
Db 1305 PCKTVNDILSTDRTKQVQCAKTKININESTPDMTCVANSKENNEYCGMCPFRKOKIC 1364
QY 1316 VHTL-----ANDNEIKLOQOVNLKEAFYSAAEFFESWYYSKDGEGNEID-KELEKSG 1370
Db 1365 LLYLKEIGENDDBOK-----FKDAFIKYAAEFELSMQYKSK---NSMDIKKLOSG 1413
QY 1371 KIPPAFLRSMEYTFEGYRDLFTGDTISKGHBESKLEQDLSLFKNGDQSPNGKTRQEM 1430
Db 1414 EIEPEELRSMTFTYGDYRDLCLNTDTSKREBDVSDAKGIDAYFNKYTDTN-----RTKW 1468
QY 1431 WTEHSHEIWEAMCAL---VKIGAKKDDFTENYNNVNFSDKSTLLEEFAPKPOFLRWL 1487
Db 1469 WDTNGBEIWEAMCALTHGVTNTDNKRKIKTQYSYKELOSKVNTTLEKFAERPOFLRWL 1528
QY 1488 TEMYDYCYCTRQYLYKDVQEK-----SNDOLKCD-TECNCKCEDYVAYMKKKKEM 1538
Db 1529 IEMGDEFCBQKRYKELKCKCNNGNVTSDCKTCVCEOKKEGKFGFI---TEW 1585
QY 1539 IPQDKYKBERDKRRDRQIHGMVTDYGTNAVDYLNKFTFASGCDKRGSAVQORNQ 1598
Db 1586 --QENMKN---QKNKEETLYQVKSRSSTISSDYIEKTL----- 1621
QY 1599 LLEKQAYYADAHGCGCTKT-ENDDKYTNISSKDKCKGLVKEANTGALIMQKGNPNYNN 1657
Db 1622 -----LKYLINEKDPYGNSTYKRAAGYTK----- 1647
QY 1658 LKEITFEDVLPFRRLRICPHALDGNTPDEVNDENGLRRLMEVAATBEGYNLGOYKKEK 1717
Db 1648 ----- 1650
QY 1718 EKEKITSDAHYSYVEPSCAMKSYFDLARDIILGIDNLEDEKOTEENLKIENKNGT 1777
Db 1651 -----IECNAVSKQNNF----- 1663
QY 1778 SVGKGSDDTTGNPGSTARKFENNEKECVWNAITGIRGRDDGNSGANSDEDLKCG 1837
Db 1664 -----ENK-----NGN----- 1670
QY 1838 SVPSDDDYPMGNKRDGTAAYOFLRWFALMGEDFCHEKKELEKLVGACNDYTCGN----- 1893
Db 1671 -----DEKAF-----KEPKETEK---ACK---CNENPPS 1696
QY 1894 -----EKRKKCDTACQYKFISEMRYQEKQIKYGENKDKLYSEHPYAKDAEDARE 1947
Db 1697 PPEELPGPATDTSYVCEVTSALT----- 1721
QY 1948 YLDKOLKICEKSGD---CEYKCKMDVSTQRLTDGNSQNPASIDDEPKEVEKCGQY 2004
Db 1722 ---IENLKACPTKYSKAPTSMKCP--SGEKSGAGATERSRADGAPSSSGSGSICV 1777
QY 2005 PRGPVRRETPSPRVSILSKATASKKAKTAPPTKQPKYENLTTEKRAOTRTTBRRAQO 2064
Db 1778 ---PRRRRLVYGLQEAETQSSQAGSSATQOEGQVYKNN-KDGASSSSSSNSQYOT 1833
QY 2065 TRKRTSTFTTESDVGIVKAILNSKPDSSRGIEGCPNKTYGYQPKWGCIVGSKENENG 2124
Db 1834 TLNASSSTSTES----- 1846
QY 2125 ICMPPRRKLCINNIOYLVNTEKRDNDIKAFIKALIFTOPLMKY----- 2173
Db 1847 -----SOLLQAFIQAIALETFLMLHRYKKEKERR 1880

```

```

QY 2174 -----IENPAENELQNGTITDEKRLMYTYGYDKUMFEGT 2211
Db 1881 RKKKKQHOHSTGTGLGLSVDEEEQPPQEBLOSGKIPYGLRQMFYLDYADKILYSG 1940
QY 2212 DISNDRKII-----TVNSVTTIENENKKQ-----DKKNOELRKIFW 2251
Db 1941 SMDNLKHIVLASGTREKQGMQKQDKIKKTLNDNNDNDESHPPSPSKNSVTTPTQW 2000
QY 2252 EKNKFIWGMITGLYHLTD-----ENEKIRIDNYQYNDMTKL 2291
Db 2001 NHAPSIWMGMRALTHKSDATYIAGGKIEQNTOLKNALLDYTKNRPKNYKY-DYVKL 2059
QY 2292 -----TPLEEFVRRPQFLRFTFEMAREFCRRKQOLKLEAGCK----- 2331
Db 2060 DENSGTSPRPAVNNAPRKLTEFERPPYFLWLEWGSFCKRQKHLEIIRVDCRGEDE 2119
QY 2332 EYECNGSN-----DEKTOCAEACVATYQNFITKMKTEYEROREKPKDKD 2376
Db 2120 DKHCSGYGENDCNLADPSIRPDLNCPGCAKHCSYKWKIQRKIEFTQODNAYNN--- 2176
QY 2377 GKXYDYSTENDIKATCAHAYLMLKLELQGNKDCSCMOKPSQLPRTYQOOSQSDAN 2436
Db 2177 -----QKVNCKE-----SKGGNGVCGLKLE----- 2197
QY 2437 DMPESLDYVPEEFNKCCEPELSKSGSMITHKTKTEPKIIMNCVKAAYLSKE---AENN 2493
Db 2198 -----ITLKEKFIPIESTEKESKNSMTNNPCDPKPYAPDKYIGRRNRCENRENR 2548
QY 2494 MD-----ITLKEKFIPIESTEKESKNSMTNNPCDPKPYAPDKYIGRRNRCENRENR 2548
Db 2215 KDNGDGTINFKEPDYTFK-----PADNCK-----PC-----SE 2242
QY 2549 FKVDYEMCYKNSKFEYOKRRVCPPRREHMCNLNDEIKTERLKD---SNYLLKAV--- 2602
Db 2243 FKID-----CTAKCTGDEBKRC-----NVANETVIRATDITKDKNGENENINNAVSD 2290
QY 2603 -----RRTANEGIDITIKNFNSE-----NCCAMPIDCMYKSYFADLGD-----IV 2643
Db 2291 TSKGDDODDLHYCHDAG--IFKGIKRDWICGNVCGV--YCKPYKVN--GQSGDNOQIIII 2346
QY 2644 RGTMLRIGYLR---PVEIKLYK-----VEYITYGK---WRN----- 2675
Db 2347 KAFKRIWIAFLIEDYNNKIKKLCSTSSDAPICIGCALWMLKKTKTEWKNLKNLYLOQ 2406
QY 2676 -KNGRKNYNDVQTFRSAMWDANRRKDIKAM-TCKAPEDAKLF-----RGRMDGFE 2725
Db 2407 YENSSDSKFLVKTILIEFND--RPFQNAIKPKCELEKFESEFGLNGADNSKSKDKGKE 2464
QY 2726 R---ITLIQ-----DYCGHKD-----DPPV--DDYIPQRFWMTEMSYYC 2761
Db 2465 RDLVLCILIEKLEKAEKCAENDADQNGVQCTQTTTDPLEDEDL----- 2509
QY 2762 KALMEELEKFKSCDHCKTSDRCNDYDENKCE 2794
Db 2510 --LLEEBENTYKAFEGCEKMEETKDKGEBCCK 2540

```

```

RESULT 12
Q8T6LO
ID Q8T6LO PRELIMINARY; PRT: 510 AA.
AC Q8T6LO:
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;

```


* Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy. "

RT J. Infect. Dis. 0:0-0(2002).
DR EMBL: AF458886; AAL84269.1; -
FT NON_TER 1 1
FT NON_TER 510 510
SQ SEQUENCE 510 AA: 60713 MW: E3CF55DC04601099 CRC64;

Query Match 14.5%; Score 2817; DB 5; Length 510;
Best Local Similarity 98.2%; Pred. No. 9, 8e-129;
Matches 501; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 233 KKHHDGMEDEYKNDYNDGSGNTYKLRAMWVNRKRWYEAITCDASTKSGTFMQSSNT 292
DB 1 KKHHDGMEDEYKNDYNDGSGNTYKLRAMWVNRKRWYEAITCDASTKSGTFMQSSNT 60
QY 293 PLPSNPGCKGKRGKVPNLNLYPOYLRFPEWEGEPCRRKRNILKRYKDCSRNDKRLYC 352
DB 61 PLPSNPGCKGKRGKVPNLNLYPOYLRFPEWEGEPCRRKRNILKRYKDCSRNDKRLYC 120
QY 353 SHNGHDCPTTIRKKGILHLDNKCTDCTCKVFEVWLGNOQEAFFKQKREYEREIOSYLS 412
DB 121 SHNGHDCPTTIRKKGILHLDNKCTDCTCKVFEVWLGNOQEAFFKQKREYEREIOSYLS 180
QY 413 NDKKFWNNISSEYKQFYEKLQYATNDTFLNLNEGKCKGGLPGEODITFTNSAD 472
DB 181 NDKKFWNNISSEYKQFYEKLQYATNDTFLNLNEGKCKGGLPGEODITFTNSAD 240
QY 473 KGIFFSEYQVCPDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLSGNE 532
DB 241 KGIFFSEYQVCPDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLSGNE 300
QY 533 QGDITOKLENCSSSTYKDKNNQKWCYKDEININCKLEONTEINNDPKIISFHNFF 592
DB 301 QGDITOKLENCSSSTYKDKNNQKWCYKDEININCKLEONTEINNDPKIISFHNFF 360
QY 593 ELWVYTYLLRDTIRKNDKLCINNTTTCIDECNRNCLCFDRWYKQKEEEMNSIKKLF 652
DB 361 ELWVYTYLLRDTIRKNDKLCINNTTTCIDECNRNCLCFDRWYKQKEEEMNSIKKLF 420
QY 653 KKNIOOSYYSNNLFEGYEFKVMKLDKDEAKKELMEINIRKKNFESLNENRDYLN 712
DB 421 KKNIOOSYYSNNLFEGYEFKVMKLDKDEAKKELMEINIRKKNFESLNENRDYLN 480
QY 713 AIELLDHLKETATICKDNNTNEACETSHN 742
DB 481 AIELLDHLKETATICKDNNTNEACETSHN 510

SULT 13
Q26033
ID Q26033 PRELIMINARY; PRT; 2664 AA.
AC Q26033;
DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR-2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; Pubmed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,
Peterson D.S., Ravetch J.A., Wellem T.E.,
"The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium falciparum-
infected erythrocytes.";
RL Cell 82:89-100(1995).
DR EMBL: I40609; AAA75398.1; -
InterPro: IPR004258; PFEMP.

DR Pfam: PF03011; PFEMP; 2.
SQ SEQUENCE 2664 AA: 302410 MW: 6EA2468511703091 CRC64;

Query Match 14.0%; Score 2714.5; DB 5; Length 2664;
Best Local Similarity 23.9%; Pred. No. 7, 1e-123;
Matches 903; Conservative 395; Mismatches 879; Indels 1605; Gaps 123;

QY 33 KSARNVLEARYAKNIRHPSYAKEHVS-----LKGLTKAEFRGSPSPVNNKHNYYR 85
DB 17 RSKAHILDSIGKKV-----YDKVHGDAIQPSNGKLTSLIAFEKAPBEQKQTSSED----- 67
QY 86 YPCNLDKHEHTNLR--YDVNLRHPCGREGONRDEDESEDCG-NKIRATKRAKDAIACA 142
DB 68 -PCDNLNHEHTTYSQDKEN--PCDKRDEVRSYTEGAECCKSKIRGSNSKDG-ACA 122
QY 143 PPRRHMCNKLAL-NDITONIHDLGVNLTAKYEGESI-----VNNHPKGT 193
DB 123 PFRHLCDQHLHRIKDKTR--HNLADYCEAKFEASLEKRYRQYQLN-----S 174
QY 194 DA---CTALARSFADIGDIVRGIDMKPN--VHKVETGLREYFKKIHGDMED-EVKND 246
DB 175 DVNINICTELARSPADIGDIVRGIDLYRGNDKEKRLLEENLRKIFKKIYDNLDAHQEH 234
QY 247 YNPD--GSGNYTKLRAMWVNRKRWYEAITCDASTKSGTFMQSSNTPLPSNPGCKHQ 304
DB 235 YKDDKGTKNYKLRAMWVNRKRWYEAITCDASTKSGTFMQSSNTPLPSNPGCKHQ 292
QY 305 KGYPTNLDYPOYLRFPEWEGEPCRRKRNILKRYKDCSRNDKRLYC SHNGHDCPTTIR 364
DB 293 NDPTTYDYPOYLRFPEWEGEPCRRKRNILKRYKDCSRNDKRLYC SHNGHDCPTTIR 352
QY 365 KKGILHLDNKCTDCTCKVFEVWLGNOQEAFFKQKREYEREIOSYLSNDKNFNANNINSE 424
DB 353 VIGHVIGSCSKSCYCRKRWIDNQKEFLKQKKCEMELSKKKOSTFYVYEG 412
QY 425 YKQFYEKLQYATNDTFLNLNEGKCK-----GGLPGEODITFTNS 469
DB 413 YKQFYEKLQYATNDTFLNLNEGKCK-----GGLPGEODITFTNS 472
QY 470 ADDKGIFFSEYQVCPDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLS 529
DB 473 NNSKRTSHSQYCECGGCGVELIGNEM--KENKKGCKGKRYNIDKTHNIVPLVSF 530
QY 530 GNEGDTOKLENCSSSTYKDKNNQKWCYKDEININCKLEONTEINNDPKIISF 588
DB 531 GDEHKEITELKIQCAENSDDSKLTQWMCYCYDKRYEYCTLENRKNSEDPBIOKTF 590
QY 589 HNFELWVYTYLLRDTIRKNDKLCINNTTTCIDECNRNCLCFDRWYKQKEEEMNSIK 647
DB 591 HNFELWVYTYLLRDTIRKNDKLCINNTTTCIDECNRNCLCFDRWYKQKEEEMNSIK 650
QY 648 KLFYKKNIOOSYYSNNLFEGYEFKVMKLDKDEAKKELMEINIRKKNFESLNENR 707
DB 651 KHFR-----YDGSITGN--NYDFVLENVLD-----ELRQDTEAENSQKIGIK 697
QY 708 DYL-----ENAILLDHLKETATICKDNNTNEACETSH-NATNPKYKPR 752
DB 698 DTLAKKTKQAADATQKNTIDLLFEYDESEAEKCK--KIQEQCPKPTKVRNPKY--- 752
QY 753 GGIQPTNINKEIQAFKRSAYEEA-RNRGLHLKLGKAHEGILKRGKGRKPFKDLCLM 811
DB 753 GNNTYDALAGVAAKIQOEAKKEQLDRNDSRLAKANSQGYTSNQGPDDFKKMLG 812
QY 812 KHSNRNLGFSNGPCDGCYKCDGIRYTHKSDNDRVNNEDYKPPMGVKNPTNITVLS 871
DB 813 KHSNA-IGDSKNPCNNNG-----KERTNVEKMWNGE-YKMSITDLTLPPRRHFCSTN 865
QY 872 LEHLQTDHPLNGNIVDLVNNSEFLGVLLSAKYEANKIIRMYEKKNLKGPEKVTDPKH 931
DB 866 LEHLNTR-----STGLTSDKAHSLIGVLLAAK-----KEGHDIK--TLTENDN 909
QY 932 OTTCRAIRYSFADIGDIVRGIDLYRGNDKEKRLLEENLRKIFKKIYDNLDAHQEH 990

Db 910 RSSICRTMKYSPADIGITINGTDLMDINGATGVONNLKIDFISKITEBELKOHDPKEND 969
QY 991 ----DABKYLKRENNWMEANRAKAYWEAMKCDIKYLKDKSGHSTQSSGYSHTPLDY 1046
Db 970 DKYINDSKHTKLRSDMWEANRDYWKMTCPTR-----NGNIQCG---ATPHDDY 1016
QY 1047 IPOKLRMTWAEYCYVQKKEKEDKKECKECKDKNDGCGCTKESGTGCTKEACNEY 1106
Db 1017 IPOKLRMTWAEYCYVQKKEKEDKKECKECKDKNDGCGCTKESGTGCTKEACNEY 1106
QY 1107 NDITLKEQMNILSDRYKELHEOQMSVNSGIEASTKKNHIDRWITELSELVQNG 1166
Db 1070 KTIQOPADQMNELSNKYQLLYMQAKIALINGSTEXTTKDKDKNVIDLQKLEHANT 1129
QY 1167 GKSNGSTDES-----AVIGTMYENGAYLHDGTFDCCOSONEFCDESKDGKD 1218
Db 1130 G-----TRPPEAHDRDRPRRAATSKSDVETTAGYIHQEARTECIGQVFCNNNG---- 1182
QY 1219 NEKAFRDRPOHDHAGCGSSKPTRVQIKTKKKAEBKDECKTYNDILKENDGKKOYE 1278
Db 1183 NMETASLPLPHEKMKACC-----NENKASS-----PEELGSDSF 1218
QY 1279 DCHPRKNSNGYDPMOCGNINLVDPVCMPPRROKLCVHFILANDNEIKKLQSGVNLKEAF 1338
Db 1219 DDH-----QTPREDEBH----- 1232
QY 1339 IKSAAETFPSTWYYSKSGEGNELDKELKEGKIPALTRSMFTYTGDRDLFTGDISK 1398
Db 1233 -----SSEGEDESEDEKE----- 1248
QY 1399 GHGEGSKLEQIDSLFKNDOKSPNGKTRQEWMTESHEIWEAMICALYKIGAKKDTE 1458
Db 1249 -----EEVEVHDGADKA----- 1262
QY 1459 NGYNNVKSSTSTLEEFKAPQFLMLTMYDYCYTRKYLKDYCKCKSNDQKCD 1518
Db 1263 ----- 1262
QY 1519 TBCNKKCEDYKMKKKKEMIPQDKYKDERDKKRDROHIGVMYTDYGTNATDYLNK 1578
Db 1263 ----- 1262
QY 1579 FTASCGDKPSASVVOIRIOLLEKQAYYDADHCGCTKFIENDKYTNISSDKCKGLVK 1638
Db 1263 ----- 1262
QY 1639 EANTGAIKQKQNNYNNLAKELTEDVLEPSRRLICFHALDGYNTDPEVKDENGILRKL 1698
Db 1263 -----GAVSQPEASPT-----TKDVV----- 1278
QY 1699 MEVATEGYNLQGYKKEKKEKIKTSAHKYSYEVPPCSAMKSYFDLRODILIGIDNLE 1758
Db 1279 -----KPPCDIYK-----ELFSSNVDLQ----- 1296
QY 1759 DEKQTEENLKIFNKNGTSVGKSGDSTTGPNGSTARKFEMNEKCVNANAMICGYKGR 1818
Db 1297 ----- 1296
QY 1819 DDGNGSNAKSPEDLKCGSVPSDDYPMGKNRDEGTAYOFLRMFAEWGDEFCHEKEL 1878
Db 1297 -----KACST-----KYGPKN----- 1308
QY 1879 EKLVCANDYTCGDNEDRKKCTOACIYKFFISEMKPOYEQKIKKYGENDKIYSEHPV 1938
Db 1309 ----- 1310
QY 1939 AKDADAREYLDKOLKIKCENKSGCEYKCMKDVSTORLTDGNSONMPASLDEPKEVEG 1998
Db 1311 -----WRCJPTKSNVYTEDG----- 1328
QY 1999 KQNCQVPRPRVRETPSPKVSLSKATASKKEAKTAPPKOKPYKVENTLTEMBAQTRT 2058
Db 1329 -----GSRHVASTPE-----SGSNSDKNGATCIPRRRLRYVGL----- 1364

QY 2059 RRAAOQTRKRTSATTTESDVGTHVAKALLSNKPDSSRGIEGCPKPYGOYPRWGCIVGKS 2118
Db 1365 -----EOWANKHNTETVSOGAEAT-----EARG-----SEAPAPG-----G 1395
QY 2119 KENENGICMPRRKRLCINNIOYLANETENKRDNDKEAFICAMIEOFLMKYIE-- 2176
Db 1396 KESSSG-----GKETPSDK-----LRTAFISAAVEFFELMDRKKEL 1434
QY 2177 -----NPAENELONGTIPDEKFRIMTYGDKDMFEFTDI 2213
Db 1435 AOKKAELONGDILYSSGDGDPNDP--QNKLLNGVLPPEFLRMFLMYLGYDRIIVHGN 1492
QY 2214 SMDKLIITYNSVTIILNENKKNKODKKDELRKI-----FEK 2253
Db 1493 TDSGNTGNSNNNNYLEASGKEDMOKIOEKLIOPLRNGSTPLVPKSSAQTPDKMNE 1552
QY 2254 NKKFWEGMIYGLY-----HLDENEKKEKIDNY----- 2283
Db 1553 HAEIWKGMICALITYEKNPDTASARGENKIEKDEVEKFFRGSTADKHGASTPTGYK 1612
QY 2284 -OYN-DMTKL-----TPSELYKRPQFLRMETAEBCNKRKQLE 2327
Db 1613 TOYDEKVKLEDTSGAKTPSASDTPPLSDFYLRPRPYRYLEWQNPCKERRKRLQIK 1672
QY 2328 AGC-----KEYECNG-----SNDGKTQCAEACTYONFIFKMKTEYEROREK 2370
Db 1673 EECMGSGDKKYSGDDEQCDRDTSNESVADLEGRSGNSCFRYKWKIRKREYDQANA 1732
QY 2371 FKDKDGKKYKDYDPTERDIEKATCAHEYLNMKLKELGKNDGSCOMKPSOLPRTQOS 2430
Db 1733 YSKOR--TKYE-----EGS 1744
QY 2431 QSSDANDPESLDYVPEEFNKECEPELSKSGSMITHKITEPKIPMCVCAAY-- 2485
Db 1745 KGAGLND-----HNKEFCVKLG-----TCTDAAPFLNRLKN 1775
QY 2486 -LSKEABN-NMDITL--KKEFLPIESTKEKESKSWTNNPCDPKPPAPADYIGRN 2539
Db 1776 GPCKDNENGGNDINFNGTEETFEFRAENCK-----PCSSFR-----INCRN 1816
QY 2540 -PCENRENNRKYVDEMYKCYNSKRYOEKKRYCVPRRRHHCLRNLEIKIERLKSNTL 2598
Db 1817 GNCRSGDG-----DTKEKC----- 1830
QY 2599 LKMYRTARNEGIDILIKFNSENGCAAMPICDYMKYSFADLGIVRGDMLRIGIYLPV 2658
Db 1831 -----NGTITTGNTGNTGCT-----EDVV----- 1851
QY 2659 EIKLYKVEYLYGWRNKNKGRNRYNDVOTFRSAMWDANRDKAMTCKAPEDAKLERK 2718
Db 1852 -----MHVSDKANNEFEGBGLDEACENAGIFGIRDEK--CKRYGSLHICQ 1898
QY 2719 GRMDGF--ERITLIQDKCGHKDPPVDYIPQFRMTKTESEYCCALMBELEKFKKSC 2775
Db 1899 EKGNGALNDQOITLVR-----ALIKRWVEYF-----LEDYKIKIKKL 1935
QY 2776 DHCTSDRCKNDYDENKCEQCTRCOEYKKNFYLKWSLFIQSKYKELY-- 2825
Db 1936 KPC-----JENGSGTICNGKKNKRNVCWIKLAKDEWTKLNNHLEKKNKBDKN 1986
QY 2826 -----EOPITYKISTYD--HYQNFVOKLKT--FKSECSVESFSEYLIHETSCLNKFN 2874
Db 1987 VTSLVTVNLETVLOIAANDKREOTSLDKLKTSLGNCPCENS-----R 2030
QY 2875 ENOSSNIRIYAFEBETPKSYKEACSCILPSPKNPLDNCPTQONKDCKELOTFPPGSKNDY 2934
Db 2031 KNDGNER-----DALDCML-----NKLETKIHECKTOH 2058
QY 2935 DNNDLNNNAVLYNSSDP--NKGVLIPRRRHLCRPIYAYVYRKGDRIILKKLITSA 2991
Db 2059 ENSVEN-----SDQPHRNGGNNPP----- 2078

```

QY 2992 FSQGLGQYKKESEELCEPAMKYADYSDIIKGTMDMTSLSEKIKITFEPSNEATEN 3051
DB 2079 -----DEEDLLEE----- 2087
QY 3052 RKTWMENNRQIWMHMLCGKIKATSKVTLDGECQOLPKDETNOPLRWLLEMAKQACKER 3111
DB 2088 -----ENPVEQ-----PGCPPTQOPEPD----- 2107
QY 3112 KHVSLSLTKTSPRSNEDNEFASELLROPCGONDIRKYYISLILIKMTMENTILIKYQOLK 3171
DB 2108 -----DKGKLEKKEKKE-----Q----- 2124
QY 3172 OSSGNIDNKPSEENVOSYIKSKDSOCALDELINDINEIVTGTNNENNEKEVLEKLYGLY 3231
DB 2125 -----EQPAEDGGAIVPSP-----PG----- 2142
QY 3232 FVEDETHKNHVLBNIKKEEQVTRPKALYFETPHVDSFYQAPLFSTHRVAQDPKNDILK 3291
DB 2143 -----SEPEADKGVKPAET-----PKP-----QEPDLSHPA-----VIP 2173
QY 3292 SSISVIVSALGL-----IALHFKKKRKS SV-DLRLINIPQEGYGMPTLESKNRYIYR 3346
DB 2174 SLVSTSLAMSVGIGFAFYFYLYLTKTKSSVGNLFQIOLPKSDYDIPKLSNRYIYPT 2233
QY 3347 SGYKRGKTYIYMGDT-----SGDEDKYMWDLSSDITSSSEYBELINDIYVPGSPRYKT 3403
DB 2234 SGYKRGKRYIYLEGGDSGYTDHY-----SDITSSSEYBELINDIYVPGSPRYKT 2287
QY 3404 LIEVLEPSPKRDIPSD-----DTPS-----NDPRTNRFIDDEENEL 3440
DB 2288 LIEVLEPSPGNNTTASGNNTTASGNNTTASGNNTTASGNNTTASGNNTTASGNNTTASGNNT 2347
QY 3441 KHDVSOYL-----PTEPNNNYKSAIDIPMTEPTLYSDNEEKPFIISIDRLYTGKEI 3497
DB 2348 KDEFISQYLOSEPTEP-----NMLGYNDNTHPTSH-HNVEEKPFISIDRLYFSEEX 2404
QY 3498 SY-----NINNSTN-----TNDIPNARNDSYIGDILINDSLVYMLLY--M 3539
DB 2405 NYMFGNSGNPNINISDTNSMDSLTSNHSPPYNDKDLVSGIDLINDALSGNHIDYDEM 2464
QY 3540 MK 3541
DB 2465 LK 2466

```

RESULT 14
P90580 PRELIMINARY; PRT: 2647 AA.

P90580: 01-MAY-1997 (Tremblrel. 03, Created)
01-MAY-1997 (Tremblrel. 03, Last sequence update)
01-DEC-2001 (Tremblrel. 19, Last annotation update)
FCR3-VART11-1 protein (Fragment).
FCR3-VART11-1.
OS Plasmodium falciplarum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
ON NCBI_Taxid=5835;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FCR3;
RX MEDLINE-97154495; PubMed-9001213;
RA Hernandez-Rivas R., Mattei D., Sterkers Y., Peterson D.S.,
Wellens T.E., Scherf A.;
RT "Expressed var genes are found in Plasmodium falciplarum subtelomeric
regions";
RL MOL. Cell. Biol. 17:604-611(1997).
DR EMBL: U67959; AAC47438.1;
DR InterPro: IPR004258; Pfam:
Pfam: PF03011; Pfam: 2.
FT NON_CODING 2647 AA; 303263 MM; 65FD700611348BC7 CRC64;
SQ SEQUENCE 2647 AA; 303263 MM; 65FD700611348BC7 CRC64;

Best Local Similarity 23.6%; Pred. No. 1,6e-121;
Matches 889; Conservative 379; Mismatches 861; Indels 1635; Gaps 120;

```

QY 31 SHKSAHVLERYAKNIRHPSYAKKEHVS-----LKGLTKEPFGG-----PSPYVK 79
DB 9 TNSAKEVLDIEIGETIQ-----KHAADATPFSQKNGEAKFTNGEITMOPNSKL-- 61
QY 80 HNYYPYPCNIDHKEHTNLYDDVNLNHPCHGREONRFEDESEEC-GNIRYKRRKND 138
DB 62 -----CELDHTIDTINVDGHSN--PEGQVTRFPDPDNRNOCCKNLIKD-SVDNSV 109
QY 139 IACAPRRHMCCKNLEBALDIN---TONHDLGNVLTAKEGSIYVNH--HGT 192
DB 110 GACAPRRHMLCSHNLNLESIOITNNYDSKASHNLAEVCYAKKEGSIYVNH--EQLSHHT 169
QY 193 SDACTALARSFADIGDIVRGIDMEKRPVHD-----KYETGLREVPKIH----- 236
DB 170 EGICTALARSFADIGDIVRGIDMEKRPVHD-----KYETGLREVPKIH----- 236
QY 237 DGMEDVKNNDYDGSNNYKLEBAMNNVNRKWEALTCDAYSKGYPMOSESNTPLFS 296
DB 230 NGAEEERYK-----DGSNNYKLEBAMNNVNRKWEALTCDAYSKGYPMOSESNTPLFS 284
QY 297 NPKCGHKGQVPTNLDYVPOYLWPMDEGEFCKRRIKAKYKDCR--NDRKLYCS 353
DB 285 GERKQCIDGIVPTNLDYVPOYLWPMDEGEFCKRRIKAKYKDCR--NDRKLYCS 344
QY 354 HNGHDCCTTIYMKGILHLIDKCTDSCPTKVFYEWLNOOEAFFKQEKYETQSYLSN 413
DB 345 RNYGDCCTTIYMKGILHLIDKCTDSCPTKVFYEWLNOOEAFFKQEKYETQSYLSN 404
QY 414 DNKFVNNINSEYKQYELKETQYATNDFNLNLEGGYCKGGLPEKDTFTNSADK 473
DB 405 SONSPPKNNMYEDEFGNLKKDYOSMD--FLKLINSEPTCTNIDAKSKIDFTK--DPE 461
QY 474 GIFYRSEYCOVCPDCGK--CDGIKTHKSDNDRERANNNDYRPMGVKFTNTLYLSGN 531
DB 462 EYFSHTEYCDPCWCGGKGTQADG-TMKRLYENDPQCEIKRKYRPPGVETEDVLYTGK 520
QY 532 EGGDITOKLENFENSSSTNYKDNKNNQKWEYCKDENIRCKLEONTLEINDNPKIIFSHNF 591
DB 521 ENKDIYKLEPCKTGNTEGK--NEBNWCYQYGN--DKCYLNGEELGEGK--KVKQDNE 577
QY 592 FELWVLYLDRITKMDKLTCTCINNNTTTCIDECNRNCLCFDRWVAKOKEEMNSIKKLT 651
DB 578 LMFVYVAMLDSDIEMRSKLNCLSDKCTCITTCNDQCQYDRIKWKYHMTQIKKHF 637
QY 652 KKKNIQ--QSYYSNNINFEQ--YEFKVDKLDKDEAKKELEMINIKRKNFENLENNRDY 709
DB 638 KQTFQGWGHYFVLETVLESDQFFDTTKAYGDRRELVHIOEMLOKKKEQV--LHEDASN 695
QY 710 LENAIELDLHLEKATATC-----KDNK-----TNEACETSHNATTPCYKPRGTOPTK 759
DB 696 MKTILDLHLEKAKQCLVNHKDNCPADLSSEDEEDIDQROKKAKP--SGTHIRA 754
QY 760 NIKELIOTYFKRSAYEERANGL--HKLKGAHEGITYKGGKRRKDKNLCAIMIKHSRNL 818
DB 755 LVNRYVSNMHHKRRKQRLVNGVSCKLGDAAKKEGRTYTIKLD--ICSIITDHSNAKR 813
QY 819 GFSNGPCDGGTGDGDIOTRVRVGTVEVDEPHMKKHEVYIMPRRRHICTSNLEHLOT 878
DB 814 GHTDQCKRDSKVVYVNR-----RMDDTAGFTISNTYKDIYMPKRRHFCSTNLEYLOT 868
QY 879 DHPLENGITVD---DLVNSFLGDLVLSAKYEANKIIRIYKKNLKLKPKREYTPKHOTJI 935
DB 869 NKLINGNDINGNPNINIDSLGDLVLAFAANEADRIKMYKN-----DKKDATI 919
QY 936 CRAIRYSFADIGIIRK-----DLMERGDAVYKLOGHLEYVGNITHSLGK 982
DB 920 CRAMKYSFADIGIIRK-----DLMERGDAVYKLOGHLEYVGNITHSLGK 971
QY 983 KGNDKYNDAPKYLKLRENNWEANRAKVWEMAKCDIYKLDKSGHOSSTOSYCGSDHTP 1042

```

Db 972 -----YKEDBP-YTKLBEDWNEANKRIWEAMOC-----PPNGSPFCKSYHIG 1014
 QY 1043 LDDYIPQKRLMTEWMAEMCYKQVQKREYDKLEKCECKDKDNNGGCKESGTGOTKTEA 1102
 Db 1015 LDDYIPQKRLMTEWMAEMCYKQVQKREYDKLEKCECKDKDNNGGCKESGTGOTKTEA 1069
 QY 1103 CNEANDIIGL--WKEOWNIISDKYKELHQAQMSVNSGIEASTAKNHIDRNVLEFLS 1159
 Db 1070 CKHVKIKIKNLILHKEQWDMKEIKYKLYLOAQTTANGSPDYSGLVDENEKPYVAFLE 1129
 QY 1160 ELYOONGK-----SNKSGTSDSAVIGINTYENYGAALH-DTGNEFDCCOSONEP 1209
 Db 1130 ELKRENGKIGNRDPTRPARSKRETAASVANDYSTAAGVHOEMGPMCECKQOTEP 1189
 QY 1210 CDEKSDGKNEKTAFRDKPODHGACGCKSGSKPTROYIKTKR-----KAEKO 1258
 Db 1190 C-EKTDEOYENENYTFKKNPQYKDACICNTRPPPKEDSRKRSDDSEBEVKETKYEKA 1248
 QY 1259 TE-----CKTVNDILKENDGKQVDECHPKKNSNGP-DMOC--- 1294
 Db 1249 TEDAVDTGPPAPKEATTIIVDCPIVAGVLT---KENENACPTKYGPAPTSWKICIP 1304
 QY 1295 -----GNINLYVEDPR-----VCMPPRROKICV---HELANON 1323
 Db 1305 EKTNAATGSGSGGNGALQAKAKATVESGSPYTSNGSICIPPRKRLYIOKLHMASGN 1364
 QY 1324 ELKILLOSQV-----NLKEAFIKSAAEFFSWYKYSKDGEGNELDEL 1367
 Db 1365 TVVSGAQTPQGGTSSPSGKETPPSDKLTAFIOSAAIEFFFLMDRYK---KSKALAKKE 1420
 QY 1368 KEGK-----IPPAFLSMWYTTGDRFLPGTD 1395
 Db 1421 KKKQWDSYPLSTADPHNPPSVLIAPNPNYKTCVIPPFLRQMYTTLGDVADIEFGKN 1480
 QY 1396 ---ISKHGE-----GSKLBQIDSLFRNGDQKSPNGRTQEWMTESHEIWEAMLCAL 1446
 Db 1481 DIYIDTRKNGDKDAERKKIKDAIEBVLKNADSQPPSDKRLQWQNEBHIMNGICL 1540
 QY 1447 VIKGAKKDFTEHYGNKWKFSDKSTLLEEFKAPROFLMTEWTDYCYTROKYLKDVQ 1506
 Db 1541 T----- 1541
 QY 1507 EKCKSNDQKCDTECNKCEDYVYKWKKKKEMIPQDKYKDEBRDKRPHROHIGVWYDY 1566
 Db 1542 -----YK----- 1543
 QY 1567 TGTNATDIYNKRFYASCGDKPGSASVQORNIQLEKQAYYADKHCCTKFTENDKYYTN 1626
 Db 1544 -----EKDEKGT- 1550
 QY 1627 ISSKDKCKGLVKEANTGAIKMQNKNPNNYNLKELTEDVLPFSRLRICFHADGNTYDP 1686
 Db 1551 ----- 1550
 QY 1687 EYKDENGIKRLKIMEVAATEGYNLGOYKREKKEKIKTSDAHKYSYEVPPCSAMKSYFD 1746
 Db 1551 PLKQEGIKSALMD-----EKNKKPKDQ-----KKQY----- 1577
 QY 1747 LRDILGIDNLEDEKQTEENLKKIFNNKGS---VAKGSDSTTGNSTARKFEMWEN 1802
 Db 1578 -----DKVKIDEN-----SGTSPKIVPAPRPTTTPPPSPSTSF----- 1612
 QY 1803 KECVNNAMICGYKRGDDGNGSARSDEDLKCGSYVSDDDYVMGNKRRDGTAYOFLRW 1862
 Db 1613 -----SRP-----PFRY 1620
 QY 1863 FAENGEDECKHKELEKLVGACNDYTCGDNEDKRKCT---DACTQYKFISEMKPOYE 1919
 Db 1621 LEWAETFEKREKRRLEREI---KVECOMDEBKQKCSGSGEGCEELRK----- 1665
 QY 1920 KQIKKYGKMKIKTSEHPVAKADADAREYLDKQKIKCENKSGCEYKCKMADYSTORLTD 1979
 Db 1666 -----ODYSTVDPFYCPECKGYCR----- 1684

QY 1980 GNSQNPASLIDEPKVECKCNCQVPRGPVRVRETPSPVSLISKATASKKEAKTAPPT 2039
 Db 1685 ----- 1684
 QY 2040 KQPKVENLTTEBRAQTRRRAAQOTRRKRTSTATTESDVGTWYKAILSNKPDSSNGIGEG 2099
 Db 1685 ----- 1684
 QY 2100 CNEKTYGQPKMGCIYVSKSENEENGICPPRRKILCINNIOYLANETENKRDNDIKEAFI 2159
 Db 1685 -----FKRW---IGKKKDEYD---KQK---AYNOKTARARN----- 1715
 QY 2160 KCAALFQGLMKYLIENPAENELONGTIPDEKRIWYTYGDKMFEPTDISNDKRI 2219
 Db 1716 -----NDNAF 1720
 QY 2220 ITVNSVTLIN---ENKKKQDKKDE-ELRKIFWEKNNKFIWEGMIGLYTHLDE 2273
 Db 1721 STTLDTCTAGDFLQTLNKGPCKNNDVDSGENKIF-----DE 1759
 QY 2274 NEKEKIRNDYQYNDMTKLTPLSEEFVKRPOFLRMETWAEFCNKRKEQLKLEAGCKEY 2333
 Db 1760 N-----GDTFKYT-----QYCG-----TCSLNGF 1778
 QY 2334 BCNGSNDGTOBCACAYTQONFIKKWTEYEROREKPKKDKGKXYDYSTERDIEKA 2393
 Db 1779 KONGD-----CVRTN----- 1791
 QY 2394 TCAHEYLNNKLKELCGNKDCSCMOKPSSQLPKTTOOSQSDANDMPESLDVYPEPNKE 2453
 Db 1792 TC-----NGSNRTTITLADI----- 1807
 QY 2454 CPELSKKSMTHTKITEPKIIPACVEKAAVYLSKEANN---DTLEKEFPIESTK 2509
 Db 1808 ---KNG-----GNSAEINMLVSDDJNSGNGFNDLBA-- 1835
 QY 2510 EKESKNSWTNNPCDPKPYAPDYKIGRNPCCENRENRFP--VDYEMCKYKKNKFPQEK 2567
 Db 1836 -----CKN--ANIRGIKENKWC---YFECK 1857
 QY 2568 KRCVPPRREHMLRNIDEIKIEHLDKSNYLLKAVRTARNEGIDILKNSENGCAMNP 2627
 Db 1858 SDVC----- 1861
 QY 2628 ICDIMKYSFADLGDIVGTDMLRIGYLPPEVETLYVFEYIYGKRNKKNKRYNDVQ 2687
 Db 1862 -----GLKKNNDI- 1869
 QY 2688 TFSAMWDANKRDIWAMTKAPEDAKLFRKGRMDGERITLLIODKCGHKKDDPPVDYIP 2747
 Db 1870 -----DONOILIRAL-----FKR----- 1883
 QY 2748 QRFWMTEWSEYCYKALMELEKFKKSCDCKTSDCKNDYDENKCEOCKTCQOEYKNFY 2807
 Db 1884 -----WLEYP---LDYINRIKRLNCP-----INNGEKAICTNGCVE----- 1917
 QY 2808 LKMSLEFDIOSNKYKELYEQPIYTKISTYDHOVONFYOKLTFSSECSVSFSYLMETS 2867
 Db 1918 -QW-----INHKRTENTNLKSF----- 1933
 QY 2868 CLYKRENDGSSNIRTYAFETPKSYKEACSCTLPSKPNLDCPTDONKDGCKEIQTF 2927
 Db 1934 ---NEOYNGDTERNPRLRFVLDLRO---IATIDK-----GNHNGYKLYKVS 1977
 QY 2928 FCSKNYDNNLDMWNAVLYVNSSDDKGVLIIPRRHLCTRPITAYNRYKGEKILKKKL 2987
 Db 1978 KC-----NCGN-----NSQNGKEGENDVLCL 2000
 QY 2988 LTSAFQGLLGQYKSEEBLCEAMKYSYADYSIDIKGTDMNDLSLKKIKIETISNE 3047
 Db 2001 L-----QKLEKKAKEC-----KDNPEYS-----GIDPODCE 2026

```

QY 3048 ATENRRTWNNRNRQIWMHMLGCKYKATATSKVTLIDEGMCOIPKDEETNOFLRWLEMAKQA 3107
DB 2027 VSENNH-----IDDE---EQPLEEEN-----T 2045
QY 3108 CKEKKHVSDSLKTCPCRSNEDNEFASELLROPCCQNDIRKYLINLIKMTMENLIRYK 3167
DB 2046 VEHKICDDVLKKNHNNORNER-----LVKRPVLOPPLTKR 2081
QY 3168 QLKQSSGNIDNKPSEBNVOSYIKSKDSOCALENDINEITVGTKNNEPKEVLKLY 3227
DB 2082 KKKKKRKKKKK-----NODFH---PRHL 2103
QY 3228 PGLFVEDETHKHNVLGDNKE---EEQTVPRKALYFPTPHVDSFYQAPLSTIRVAYQDP 3285
DB 2104 PCGAFINTNTPKTKTPSSSGKNPWEHPAVLPA----- 2135
QY 3286 KNDLKSSISVIVSALGLALHFKMKFKSVLDLRLINIPQEGYMPLESKNRYIPY 3345
DB 2136 ---LVSTLAMSIGIGPAATFYLLAKKTKSTIDLLSL-IPKSDVDIPTKLSPNRYIPY 2191
QY 3346 RSGPYGKTYIYEGDT---SGDEKYMMDLSSDIT-SESEYEELINDIYVPGSPKY 3401
DB 2192 TSGKYRGKRYILEGSDGTSGYTDHY-----SDITSSSESEEMDINDIYVPGSPKY 2245
QY 3402 KTLIEVLEES-----KNDIPSDTTPSNDTPRTNPFIDEMWELK 3441
DB 2246 KTLIEVLEESGKLSGNTIPTSGNNTTASDPTQNDIDIPSTDPPT-----DDEWMTLK 2298
QY 3442 HDVVSQVLPTEERN--NNKYSADIPMTEBPTLYSD-----NPEKPFITSI 3486
DB 2299 HDISNMLQNPQDPVPRDYISGNSSTNTITTSRDVNDVNNHTPKSRHNVDDQKPFITSI 2358
QY 3487 HDRDLTYGKEISTYNKSTNTNDIPMANNDSTRGIDINDLSI 3530
DB 2359 HDRNLYGEEYNYNVNM-VMTMDIPINSHNNVSGIDINDLSI 2401

RESULT 15
ID 094657 PRELIMINARY; PRT; 2212 AA.
AC 094657;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE Erythrocyte membrane protein 1.
GN PEPMP1.
OS Plasmodium falciparum.
NC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;
[1]
RP SEQUENCE FROM N.A.
RA STRAIN=JTG.
RA Yang S.O., Wollish W.S., Gut J., Wu J., Ahn J., Petersen C.,
RA Fujioaka H., Aikawa M., Leach J.H., Nelson R.G.;
RT "The molecular cloning and DNA sequence analysis of Plasmodium
RT falciparum erythrocyte membrane protein 1."
RL EMBL; U31083; AAB06961.1; .
DR EMBL; U31083; AAB06961.1; .
DR InterPro; IPR004258; PEPMP.
DR Pfam; PF03011; PEPMP; 2.
SQ SEQUENCE 2212 AA; 253594 MW; A33456648C852FC CXC64;

Query Match 13.8%; Score 2677; DB 5; Length 2212;
Best local Similarity 31.1%; Pred. No. 3.7e-121;
Matches 713; Conservative 307; Mismatches 800; Indels 476; Gaps 83;
QY 31 SHSASRVLERYAKNRHPSKYAKEHVS-----LKGLDITKAEFRGG-----PSTVYNK 79
DB 9 TNSAKVYLDIGETIIO-----KKAHSDADTPSQSLKNGEAKFYNGEITQMOVNSKL-- 61
QY 80 HNYVYPCMLDKRHTNLRKYDVNLRHPCGHEQNRNFEDESEEC-GNIRYKRRKND 138
DB 62 -----CELDHTIDINVTGHSN---PCGQTVAFRPDNRNSQCKKNIKD-SVDNSV 109

```

```

QY 139 IACAPRRRRMDCKNLEALNDIN---TONIHDLGNLVTAKEGESIVNNHP---HKG 192
DB 110 GACAPYRRRLHCSHNLESISTQNNYDSSKAKHNLAECYAAKEGESIVNVEYOLGHTT 169
QY 193 SDACTALARSFADIGDVRGIDMKRPVND-----KVEFLREVEKTIH----- 236
DB 170 EGICTALARSFADIGDVRGIDMKRPVND-----KVEFLREVEKTIH----- 229
QY 237 DGMEDVEYKNDYNDPDSGNYKLRAMANNVNRNKAWEAITDASYSKYGFMOSESNTPLFS 296
DB 230 NGAEBERYK-----DGSNNYKLRDMWANNRBLDWMKMICAKAPNOAQYFRNTCSNGEPT 284
QY 297 NPKCGHOGVPPNLDVYVPOYLRFEDWGEFECKRRIKATKAYDSR---NDKERLYCS 353
DB 285 GERKQCIDGVPPNLDVYVPOYLRFEDWGEFECKRRIKATKAYDSR---NDKERLYCS 344
QY 354 HNGHIDCTTITMKGILHLDNCKDCTSCYKCYEVEWLNQOEAFAKOKEKYKETSQSLN 413
DB 345 RKGIDCTKTRISDKYSMNRECTKLYVCDPYKWMIDNKKKEFEKOKKCENEIYRNNE 404
QY 414 DNKFVNNINSEYKYQFYEKLKETOYATNDFNLNLENGYCKGGLPEKDIPTFNSADK 473
DB 405 SQNSPKYNNMYETDFGNLAKVOQSMND-FLKLNSETCTNIDKKRIDFTK--DPE 461
QY 474 GIFYRSEYCOVCPDQCYK--CDGIKTYHKSNDREVRNNEDEYKPPWCVKPTNITVYSGN 531
DB 462 ETSHTLEYCDPCPWCGLKQADG-TWRRLYENDPQCIFKPEYEPKEEPTFIDVLYTGK 520
QY 532 EOGDITQKLENFCSSTNYNDKNNQKWEYKEDINRKNLEONTETLNPNKILSHNF 591
DB 521 ENKDIYKLEFCCTGDNTEGK-NEENNYCYOYN-DYVLENGEELGEGK-KKQDNE 577
QY 592 FELWVLYLDITKMDNKLKTCJINTTTHCIDECNRCLCFDRVYKOKEBEWSNIRKLYT 651
DB 578 IMFVVAHMLKDSIEMRSKLSNCLSDKRTCKIKKNDCKCYEKVIGKKVYEMTQIKKHF 637
QY 652 KKKRNIQ-QSYTSNINNFEG-YFFKVDKLDKDAKKELEMNIRKKNKNEFSNENRDY 709
DB 638 KQTFQOQMGFRFVLEYLEGDQFPTDITKAYGDAARELVHIOEMOKKREOV--LHEASN 695
QY 710 LENAIELLDHLKRETTATC-----KDNNTNEACETSHN-----ATPPCVPRGCTOPTK 759
DB 696 MKTIIDELDLHELKAEACQVYNHNDNCPDLSSEDEEEDIPORONPCAKPSSABRAL 755
QY 760 NIKELIAYPKRSAYEARNRGL-HLKGAHEGIYKRGRRKDKMLCRIMKHSNRL 818
DB 756 -VNVASNMHKKKRLQVNVGSVKLKGDAKGEYRSSTIKLKD-ICSTINDHNAKR 813
QY 819 GFSNGPCDGGKGTGDTOTRFVYGTIEWEYDEPHMKKHEDVIMPPRRHICTSNLEHLOT 878
DB 814 GHDPQPCRRK---DSKSEMFTEDEGWR-PAGFISKTYKDIYMPRRHFCYSNLEYLOT 869
QY 879 DHPNGNIYD---DLYVNSFLGDVLLSAKYEANKIIMYKREKNLKLKPKETDPRKHOTTI 935
DB 870 NKLNGNDINGNPNVINDISFLGDVLPANTEADRIKMTKKNQ-----DYKDANTI 920
QY 936 CRAIRYSFADIGDITIRGRDLMERNGDWKILQGHLETFYGNHKSILKAGKNDKYNDAPKY 995
DB 921 CRANKYFADLGDIITKTDMDNDNSGSKTRDRELFTDIKKKHPI-KEIYEDPR-Y 978
QY 996 LKLENNWEANRAKAWEMAKCDIYLLKDSGHSTQSSYCGYSDHTPLDYITPOKLRMT 1055
DB 979 TKLREDMWEANRRKIWEAMOC-----PPNCSFPCKSYHTPLDYIIPORLWMT 1027
QY 1056 EMAEYWCYKQKKEVDLKECKECKDKDNQGGCKREGSGTGTCTEACNEENDIIGLAK 1115
DB 1028 EMAEYWCYKQKKEVDLKECKECKDKDNQGGCKREGSGTGTCTEACNEENDIIGLAK 1084
QY 1116 QWNITISDKYELHBOAOMVSNSGIEASTAKNIDINRVLEFISELYQONGK----- 1168
DB 1085 QWDMKEIKYLLYLAQATTAANGPDIYSGLVYDENKRPVNFLEFELKRENGGKIGNRDT 1144

```

QY 1169 --SNKSTSDSAVIGINTTYENAGAYLH-DTGNFDDCQSONEFCEKSDGKDNKAYFR 1225
 Db 1145 PRKRSTRETPASAYAKNDVYSTAGYVHOEMGPHMECKTQTEFC-EKTDQENENYTFK 1203
 QY 1226 DKPODHGACGCKSGSKTRQIYTKK-----KAEKQTE----- 1260
 Db 1204 NPPQYKDACICITNPPKEDSRKSEDSDEBEKYETKEKATEDAVDTGPPAPKEA 1263
 QY 1261 -----CTVNDILKENDGKROVEDCHPKKNSNGYP-DMOC-----CNI 1297
 Db 1264 TTTLIDVCIYAGVLT-----KENLENACPTKYGPAPYSMKCIPTKTAATGSESSGNG 1319
 QY 1298 NLVEDPR-----VCMPRRQKLCV--HFLANDNEIKKLQSOV----- 1332
 Db 1320 ALQRAKRAFVSGSPVTSNGSICIPRRRLRYIQLKHDWASGNTVSGOQOTPOGTS 1379
 QY 1333 -----NLKEAFKSAAEFPFSWYKSKOGEDELKE----- 1369
 Db 1380 PSGKETPSDKLTAFIQAIAETFFLMRYKE-----KEIEKKEKRVANGGLVPSLNGP 1435
 QY 1370 -----GKIPAFIRSMFYTFGDRDPLFTD---ISKHGE----- 1402
 Db 1436 PQOPGVYGDSPQSKIQGCVIIPPELRQMFYTLGDIADIFGKNDIYIDTKNGDKDIAR 1495
 QY 1403 GSKLEQJDSLFKNGDQKSPNGKTRQEWTEHSEIWEAMLCALY-----KI 1449
 Db 1496 EKKIKDAIERVYKKNADSQPPSPDEKQOTWMEQNGEHINMGICALYKKEDEKGTPLKONE 1555
 QY 1450 GAK-----KDPTEYGYNNVYFSDKS-----TLEEFAPKRPQFLRWLJE 1489
 Db 1556 GLKSLMDEKKNKPPDOYQYIDKYLKDNESGTSPTKNDHPPTPLNTISRPPYRYLEE 1615
 QY 1490 WYDDYCYTRQYLYKDVQEKCKSND--QLKCD-----TECNKRC 1525
 Db 1616 WGEFECRERKRLEKIKYECMEDEGKQKCGSDGECIEIRKODYSTVDFYCPGCKYK 1675
 QY 1526 EDYKYMKKK-EMIPQKYYKDERDKRFRQIHGVAVTDYTGYNADYLNKFTASC- 1583
 Db 1676 RPYKRWIEKKDEYDKQEAANNQCTDARNNNDNAFSTLTCTAGDFLOTLKNPCK 1735
 QY 1584 ----GDKPGSASVQORNIQLEKQAYYADKHCCTKFLENDK---TYNISKDKCKGL 1636
 Db 1736 NDWVDDSGENKKIPDENDGTRKYIYCGT---CSLNGFKCNGDDCRVRYNT---CNGS 1788
 QY 1637 VKEANTGAIKQWONGPNYNNMLKELTEDVLPSSRLRICFHALDGNVYTDPEYKENDGLRK 1696
 Db 1789 NRTTITADDIKNG--NSAEINMLVSD-----DINGNGFN- 1823
 QY 1697 RLMEVAATEGINLGQYKEKEKEKIKTSDAHKYSEYVPPCSAMKYSFYDLRDTIL--- 1752
 Db 1824 ---DLEACKNANIFKGIKENKWK-----CVYFCKSDVOCGLKKNNDIDONOILIRAL 1872
 QY 1753 ---GIDNDEDEKOKTEENMLKIFNKNGTSVGKSDSTGNPSTARKFFWMENKCEVWNA 1809
 Db 1873 FKRLIEFYLDIYKIRKILNFCINN---GKAICTNGCVQ-----WINOKTEWT- 1920
 QY 1810 MICYKRRGRDNGSNGSARSDEDLKCGSVPSDDYPMCKNRDEGTAYOFLMFAWGED 1869
 Db 1921 -----NIKN--RFNEQY-----NGDDTEM-KSFSRSLVLDLIRQIA---AT 1955
 QY 1870 FCKHKEKELEKLYGA--CNDTTCGDNEDKRRKCTACTOYKFFISEMKPQYKQIKKYE 1927
 Db 1956 IDKGNHNGLVKLVSVKN---CGNNSONGKE-----GE 1986
 QY 1928 NKDKIYSEHPAKDAEDAREYLDKOLKICEN--KSGDEYKCMADVSTQRLTD----- 1979
 Db 1987 ENDVLC-----LLQKLEKAKECKDNPEISGIPQPC--EVSBNHTEDEQPLE 2034
 QY 1980 --GNSQNNPASLD---EPREVECKCQYPRGPPRVARRETPSPRVSILSKATASKKEA 2033
 Db 2035 EEENTVHEPKICDDVLTQPOPEERGETCEBSPGPTDVKEEEEEK-----KEEKDEEB 2089
 QY 2034 KTAPTKQPKVENLT 2049

Db 2090 PGLPPTPPAPASST 2105
 Search completed: April 28, 2003, 10:35:24
 Job time : 233.729 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 35.7348 Seconds
(without alignments)
2916.372 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407
Sequence: 1 MGFSCKYFIIMGNASSLE.....IDLINDSLVNLILIMKY 3542

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

al number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3489	18.0	3060	US-08-487-826B-14	Sequence 14, Appl
2	2972.5	15.3	2710	US-08-568-459A-12	Sequence 12, Appl
3	2972.5	15.3	2710	US-08-487-826B-12	Sequence 12, Appl
4	2972.5	15.3	2710	US-09-210-288-12	Sequence 12, Appl
5	1743	9.0	2182	US-08-487-826B-16	Sequence 16, Appl
6	1032.5	5.3	700	US-08-568-459A-10	Sequence 10, Appl
7	1032.5	5.3	700	US-08-487-826B-10	Sequence 10, Appl
8	1032.5	5.3	700	US-09-210-288-10	Sequence 10, Appl
9	633	3.3	921	US-08-568-459A-8	Sequence 8, Appl
10	633	3.3	921	US-08-487-826B-8	Sequence 8, Appl
11	633	3.3	921	US-09-210-288-8	Sequence 8, Appl
12	617	3.2	1435	US-08-568-459A-4	Sequence 4, Appl
13	617	3.2	1435	US-08-487-826B-4	Sequence 4, Appl
14	617	3.2	1435	US-09-210-288-4	Sequence 4, Appl
15	431	2.2	749	US-08-568-459A-6	Sequence 6, Appl
16	431	2.2	749	US-08-487-826B-6	Sequence 6, Appl
17	431	2.2	749	US-09-210-288-6	Sequence 6, Appl
18	396	2.0	10182	US-09-134-001C-3159	Sequence 3159, Ap
19	381	2.0	3135	US-08-323-170B-2	Sequence 2, Appl
20	381	2.0	3135	US-08-954-441-2	Sequence 2, Appl
21	365	1.9	411	US-08-568-459A-20	Sequence 20, Appl
22	365	1.9	411	US-08-487-826B-32	Sequence 32, Appl
23	365	1.9	411	US-09-210-288-20	Sequence 20, Appl
24	361	1.9	3248	US-08-353-700-1	Sequence 1, Appl
25	361	1.9	3248	US-08-568-459A-1	Sequence 1, Appl
26	343.5	1.8	1115	US-08-568-459A-2	Sequence 2, Appl
27	343.5	1.8	1115	US-08-487-826B-2	Sequence 2, Appl

28	343.5	1.8	1115	4	US-09-210-288-2	Sequence 2, Appl
29	343.5	1.8	1115	6	5198347-6	Patent No. 5198347
30	340.5	1.8	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
31	291.5	1.5	2482	1	US-08-328-254-6	Sequence 6, Appl
32	279	1.4	362	2	US-08-568-459A-18	Sequence 18, Appl
33	279	1.4	362	2	US-08-487-826B-30	Sequence 30, Appl
34	279	1.4	362	4	US-09-210-288-18	Sequence 18, Appl
35	276.5	1.4	2285	4	US-09-308-375-2	Sequence 2, Appl
36	276	1.4	411	2	US-08-568-459A-19	Sequence 19, Appl
37	276	1.4	411	2	US-08-487-826B-31	Sequence 31, Appl
38	276	1.4	411	4	US-09-210-288-19	Sequence 19, Appl
39	273.5	1.4	311	2	US-08-568-459A-21	Sequence 21, Appl
40	273.5	1.4	311	2	US-08-487-826B-33	Sequence 33, Appl
41	273.5	1.4	311	4	US-09-210-288-21	Sequence 21, Appl
42	257	1.3	3418	3	US-08-755-587-44	Sequence 44, Appl
43	256.5	1.3	1663	5	PCR-US93-07261-16	Sequence 16, Appl
44	251	1.3	3418	2	US-08-639-501-2	Sequence 2, Appl
45	251	1.3	3418	3	US-09-044-946-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-487-826B-14
; Sequence 14, Application US/08487826B
; Patent No. 5993827
GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chittis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelisen, Ned
; REGISTRATION NUMBER: 29,655
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-826B-14
Query Match 18.0%; Score 3489; DB 2; Length 3060;
Best Local Similarity 26.7%; Pred. No. 1.5e-240;
Matches 1054; Conservative 472; Mismatches 934; Indels 1490; Gaps 148;

QY 9 IIRKGNASSLEBDGAKSPIIKESHKSARVLEYAKNIRHPSK-YAKEHYDSIKGLDTKA 67
 Db 4 LAKMGPEAGAGDIED-----ESAKHMFDRIGKVOYKAVEEKEKREKQIGRLISEA 56
 QY 68 EFRGGBSTPVANKNNYYPPYPCNDHKHENTLRDVOVLAHPCGREGREONFDEDESEBC-G 126
 Db 57 KFEKNESDPOTPEB-----PCDLHKYHTNV-----TJNVIPCADRSVPRESDEYGGQCTH 108
 QY 127 NKIRNYKRNKDALACAPRRRRHCDKNLEALDINTONIHDLGNLYVTKYGESIYVN 186
 Db 109 NRIKDSQGGNKGACAPRYRLHYCDONLREIEPILKTNTNHLVYDCAKAFEGQSTIOD 168
 QY 187 HP-HKGT-----SDACTALARSADIGDIYRGIDMEKPNVHD-----KYETGLREVEFKKI 235
 Db 169 YPKYQATYGDSPQICTMALARPADIGDIYRGHDLVIGNFOEIKORQOLENNKLTIFGKI 228
 QY 236 HD-----MEDEVKNDYNDPDSGANYKILREAMWNNRKNKWEALTCDSYSGYMOSESNT 292
 Db 229 YELKNGAEARYND--PE-----FEKLRDMWTANRRTWKALTCNA--WGNTYF-HATCNR 280
 QY 293 PLESPNKGCHROKGVPTNLDYVPOYLWPEDEWGEFRCRRNIIKIKYVKDSCR-----NDKE 348
 Db 281 GERTKGYCNCNDQVPTLYPDYVQYLRWFEEMAEDECRKKNKIKYKNCCKKDEKD 340
 QY 349 RLYCSHNGDCTTTWKKGILHDKCTDCTCKYFEVWLGNOEAFKOKKYEKEIO 408
 Db 341 R-YCSHNGYDCEKTKRAIKRLRYGKQICSLACNPNYMINOKQOFKOKKYEKEI 399
 QY 409 SY-----LSNDNRKFNVINNSE-YKOFYEKLEKTOYATNDPFLNLSGKYC-----454
 Db 400 KXENGASGSRKRDAGGTTTNYDGEKKEFDLKNSEKRYVDKLEKLSNEEICTRYK 459
 QY 455 --KGLPEKEDYFTNSA-----DDKGIYRSEYQVCPDQCGKDEIKYTH--KSDND 504
 Db 460 DEEGGTIDFKVNSSTSGASGTNVEOGTFYRSKYCPCPYGVK-----KVNNGSSSNE 515
 QY 505 REEVNT-----EDKPPGVKPTNTVYLSGNEGDIQOKLENCN-----SS 547
 Db 516 WEKNNNGKCKSGKLYEPKDKEGCTITLLKSGKHDDIEKLNKCEKDEKNGDTINSGGG 575
 QY 548 TWYKDKNN-----OKWECYKDENINRCKLEONTIENDNPK-----II-----586
 Db 576 TGGSGGSGNSROELYEEMWC-YKGEDVYVVGHEDEDEDEYENKNAAGCJILKOKKNE 634
 QY 587 -----SEHFEELWYTYLIRDIKANDKIKTI--NNTTTHC-IDECNRNC 629
 Db 635 ECGNTSEKPEDEIOKTFEPFYVWAHMLKDSIHKKKLOKLOCLONGNRILKCNKNCNDNC 694
 QY 630 LCFDRVYKQKEEEMNSIKLFTKKNKIQ-----QSYYSNINLFEGYF--672
 Db 695 ECFKMWITQKKDWEKIVQHF-KTONIKRGSGDWTALIPPDHIVYLOYNLOEFLKGD 753
 QY 673 -----FVYMDKLDKDEAKWELMENIKRKNNEFSLENND-----YLENAIELL 718
 Db 754 SEDASEESENSLDAEEA-----DELKHLREIIESEDNNOASVGGVTEGKNIMDKL 807
 QY 719 DMLKFTATTC-----KDNNTNACELSHNATINPCVYKPGGQO--PKNKIKELAO 766
 Db 808 NYEKEDADICIEIHDEEERKKGCGNCTIEGENFRYNPCSGESGNKRYPLANKVAYQ 867
 QY 767 YFKRSAYEARNRGLIKLKAHEGIRYKGRKRDKDNLCRIMKHSNRMNGSNGCD 826
 Db 868 MHHKAKQOLASAGSARLGRDLSLAQFKNGRNGSLKGOICKINENYSDNRSGSGGCT 927
 QY 827 KCGGSDGIGOTFRVVGTEWEPDEHMRKDEHVIIMPRRRHICTSNLEHLOTDHPLNGNI 886
 Db 928 GK-DGDHGGVMRIGTESNIEGKKQTSYKKNVFLPRPREHCTSNLEULV-----GSV 980
 QY 887 V-DDLVNNSPFGDVLISKYVANKIIRMYKREKNILKGEKVTDP--KHQTTICRAIYS 942
 Db 981 TKNDKASISLGDVQOLAKTKTAELIKRYKDONNI-----QULTDPOQDOEAMCAVAYS 1036
 QY 943 FADIGDIIIRGHDWERNGDVVLQGHLETVGVNIIKSLKG-KGNDKYND--APKYLKL 998

Db 1037 FADIGDIIIRGHDWERNGDKSTDETRLTIVFNKIEKHDIQNDNRYTDESCKKAYKYL 1096
 QY 999 RENWMBANRAKYWEAMKCDIKYLLKDKSGHQSOTSGYSGDHPDDYIPOLRMTTEWA 1058
 Db 1097 RADWMBANRHOYWRAMK-----ATKGIIC--PMPAPDDYIPOLRMTTEWA 1141
 QY 1059 EWCYKQKKEYOIKLEKCECKDKNQCGCTKESGCTCKTEACNEYNIDIIIGKEOWN 1118
 Db 1142 EWCYKQSOEYKLEKKTICADMSKDGK-CT-QGVNDCGCKCAADKYEEIEKWNEMWR 1199
 QY 1119 IISDKYKEHEQAOAMSNSVSGIEASTAKNHID--RNVIELSELXOONGK-----SNK 1171
 Db 1200 KISDKNMLLYLAKTSTNPG-----RTVIGDDDPYQKQVDELTYIHKASIAARYLARA 1255
 QY 1172 SGTSDSEAVIGTNTYENVGAYLHDTPGNDQOSONECDEK-----SDGKNKVAFR 1225
 Db 1256 ASPEIEMAAABITYSTAGYIHOEIGGCOEQOTCEKKGATSTSTKENNEYTFK 1315
 QY 1226 DKPOHDGACGSGSKPTRVQIKTKKAKEKDTECTYVNDILEKNDOKKOEYDCHPKN 1285
 Db 1316 QPPEIATACDINSQTE-----EPKKEENESACKIVKILEKNGRTVYGECPKES 1371
 QY 1286 SNGYEDMOC-GNINLVEDPRVCMPPRROKLYHFLANNEIKKLOSOYNLEAFITKSAA 1344
 Db 1372 ---YDWMCKNNIDISH-GACMPPRROKLYIAHSQENITDNLDAFIKIAA 1427
 QY 1345 EFFEYVYIYKSK-DOEGNDELKELKGIIPALRSMFYTGDRDLFGTDISKHGBG 1403
 Db 1428 EFTLSMOYKSKNDESEALIR-----GLIPSOFLSMAYTGDYHDICLNTDISKQNDV 1483
 QY 1404 SKLEQIDSLFKNGDQSPNGKTROEMWTEHSEHWEMLCALVIGAKKD--FTEHY 1460
 Db 1484 AKAKIKIKFISKDSKSPSGLSROEMWKTGPELWKMALALTRYVDTNKKRKNKY 1543
 QY 1461 GYNNYKESDK-STYIEEFAKRPOLWLTWYDDYCYTRORYLKDVOEKCS-NDOLJK 1518
 Db 1544 SYDKVNOSONNPSLIEEFAKRPOLWLMIEWEECAERKKEINIKDACNEINSTQCN 1603
 QY 1519 T--ECNKCEYVYKMK-KKEWIPQDKYKDERDKRFRPROHIGVATDYTGNTATDY 1574
 Db 1604 DAKHGNACARAYOYEVNKKKFEESGOTNNVLANVOPPOPEKGEYKXO-----1654
 QY 1575 LNRFTASCGRKPGSASVYVQNIQLEKQAVYDADKHCQKTFENDKYTNISSKDKK 1634
 Db 1655 -----GVOP-----IQENEYILQK--CDNNK-CSC-----1676
 QY 1635 GLVYEANTGALKMONGKGNNTNNIKELTEDEVLFPSRLRLICFHALDGNVTDPEYKDNGL 1694
 Db 1677 -----MDGNVLSVSPKEX--1689
 QY 1695 RKRLMEVATBGYNLGOYKKEKKEKIKTSDAHKYSYEVPPCSAMKYSFYDLBIIIGI 1754
 Db 1690 -----PPGKY-----1699
 QY 1755 DNLEDEKQTEENLKIFNNKNGTSVKGSDSTGNPGSTARKFWEKNEKCVNNAICGY 1814
 Db 1700 -----1699
 QY 1815 KRGDDDSNGSASASDDLLKCGSVPSDDIYPMKANDBSTANQFLWFAWGEDFCCHK 1874
 Db 1700 -----1699
 QY 1875 EKELEKLVGACNDYTCGDNEDKRRKCTDCTOYKFKFISEMKPOLYEMOIKKYEKNKDIYS 1934
 Db 1700 -----EKC-----1702
 QY 1935 EHPVAKDAEDAREYLDQOLKIKCENKSGDEYKCMKDVSTORLTGNSQNMPSLIDEPK 1994
 Db 1703 -----DC-----1704
 QY 1995 EVEKCMCOYPRGPRVRRRETSPRVSLSKATASKKEAKTAPTPQPKVENVLTTEMRA 2054

QY 1635 GLVKEANTGAIKMNKGNPNYNNLKEITEDVLPSPRRRLICFHALDQNTDTPREYKENG 1694
 Db 1679 -----MGNVLSVSPKEK----- 1691
 QY 1695 RKRLMEVAATEGYNLGOYKEKKEKIKTSDAHKSYEYRPPCSAMKYSFYDLRILIGI 1754
 Db 1692 -----PRGKY-----AKHY----- 1701
 QY 1755 DNEDEKOTEENLKIKIFNNGTSVCGKSDSTJGNGSTARKFEFNNKECVNNAICGY 1814
 Db 1702 ----- 1701
 QY 1815 KRGRDQNSGNSARSDDELKCKGSPBDDYIPMGKNRDEGTAYQFLWPAEMGEDEPKKH 1874
 Db 1702 ----- 1701
 QY 1875 EKELEKLVGACNDYTGDNEDKRRKCTDCTOYKFISEWKPQYKQIKKYENKDIYS 1934
 Db 1702 -----EKC----- 1704
 QY 1935 EHPVAKDAEDAREYLDKOLKIKCENNSGDEYKCMKDVSTORLTDGNSQNMPSLDEPK 1994
 Db 1705 -----DC----- 1706
 QY 1995 EYEGKCMQYPRGPRVRRPSPRVLISKATASKKEATAPPTKOPKVENLTTEMA 2054
 Db 1707 -YQK--HVPSP-- 1725
 QY 2055 QTRRAAQOTRRKTSTATTESDVGTMVRAILSNKPSRSGIEGCPNKTYGQY--PKWG 2112
 Db 1726 -----PEAPVTVYDVCISIVTLK--DTNFBSDACKL-YGKTAPSSWK 1766
 QY 2113 CT-----VGSKENENGICMPRRKLCINNIOYLTET-----ENKRD 2151
 Db 1767 CIPSDTKSGAGATTKGSGSDSGSICIPRRRRRLYVGKIQ--EMWATLPQEGAPSHSRA 1824
 QY 2152 NDIKEAFIKCAIETQPLMKIYIE--NPAEN-----LONGT 2188
 Db 1825 DDLENAITQSAIETFLMRYEKKRPGQSGOQALSLTSTYSDEDEPPDKLLQNGK 1884
 QY 2189 IPDEFKIMVYTYGDKMFEFGDISNDKIIYVNSVTLLNENNRKODKKDEELRK 2248
 Db 1885 IPPDFLRLMYYTIGDYDILVHGNTSDSGNTNSNNNNIYVLESGKEDMOKQOEIEQ 1944
 QY 2249 I-----FWKKNKFIWEGMITYGLY-----HLDENEKIR 2280
 Db 1945 ILPRNGSTPLVPRSSAQTPDKMWNHEASIMKMICALITYEKNPDTSARGDERKIEKD 2004
 QY 2281 DNY-----OYN-DMTKL-----TPSLEEFYKRP 2302
 Db 2005 EYEEKEFGSTADKHGTASTPTGYTKYDYEKYKLEDDTSAGKTPSASSDPPLSDFVLRP 2064
 QY 2303 QFLMFEFMAEFCNKRREOLKLEAGCKEYE-----CN---GSN 2339
 Db 2065 PYRYILEMONGNCKKKHHLAQKHCKYEBENGSGSRGIGITROYSGDEACENMLPKN 2124
 QY 2340 DGTQOE-----CAEACVLYNFIKKMTEYEROREKFRKRD----- 2376
 Db 2125 DGYPDLKEPKSCAKPCSSYKRWIESKGEFEKQEKAYEQKDKVNSGNKHNDGFCETLT 2184
 QY 2377 -GKKYKXY-----PSTRDRIEKATCAHEYLNMKIKELCKGKDKSCMKRPSOLPKTQOS 2430
 Db 2185 TSSAKQFLTLTGPCCKNNVEGKITFDODTKF-----HTKDDPCLKFSVCKR----- 2234
 QY 2431 QSSDAMPDPSLDYVPEEFKCECPPELSKSGSMITHKTIKTEPKIPMNCVEKAAAYLSKEA 2490
 Db 2235 -----DECD-----NSKGTDCRKN-----NSID-----ATDI 2266
 QY 2491 ENNDITLKEKFLPIESTKKEKESKNSWTNNPCDPKPKVADKYIGRNPCE----- 2543
 Db 2257 ENGVDSTV-----LEMRVADSCKSGFNCD-----GLENACRGAGIEGI 2295
 QY 2544 REBRFVVDYEMKCYKNSKRYOEKKRVCVPRRREHMLRLNDELKIRLDSNVLKMRV 2603

Db 2296 RKD-----EMKC-RNVCGY-----VCKP----- 2313
 QY 2604 RTARNEGIDILIKNNSENGCAMPICDTMYKSPADLADIVRGTDMLRIGGLPVEIKLY 2663
 Db 2314 -----ENVNG-----AKGHIIDIRALVR-----RWV 2336
 QY 2664 KVEYIYKWRNKNKNGKRYND-----VQFRSAMWMDARRDIKAMTKCAPEDAKLERKG 2719
 Db 2337 EYFEEDYKRIKHISRIKNGEISPCKIKNCVEKWDQKRRE-WKEITERKD----- 2387
 QY 2720 RMGFEIITLQDKCGKXKDDPPVDYIPQFRMTESEYCCALMELEKFKKSCDHCK 2779
 Db 2388 -----QYKNDSDDDNRSFLETLLIPQTTDAN---ANKRYIKISKFGNSG-CS 2432
 QY 2780 TSDRCKNDYDENCCEQCTRCOEYKNFVLWKSILFDIOSKRYELYEQPIYTKISTYDHV 2839
 Db 2433 AS---ANEQKN-----GEKDAI-----DMLKRLND-----KIGCEBK 2465
 QY 2840 QNFVQKLTFRKSECVSEFSEYLIHETSKLNYKFNENDGSSNIRTYAFETPSYKE--- 2896
 Db 2466 HH-----QTSDTRECSDPPOQTLLED--ETLDDIETEEAKKNMMPKICENVLTKAQODE 2518
 QY 2897 -ACSTLPSKKNPLDNCPTQNDK 2918
 Db 2519 GGC---VPAENSEPAAATDSKE 2538
 RESULT 3
 US-08-487-8268-12
 Sequence 12, Application US/084878268
 Patent No. 5993827
 GENERAL INFORMATION:
 APPLICANT: Sim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knodde Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelso, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO

```

; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-12

```

Query Match	15.3%;	Score 2972.5;	DB 2;	Length 2710;
Best Local Similarity	26.5%;	Pred. No. 1.3e-203;		
Matches 865;	Conservative 442;	Mismatches 873;	Indels 1083;	Gaps 120

[illegible]

Db	930	GK-DGSHGVRRIRIGTWSMINEGKKQJTSYKVVLPFRPHHXTSINLENDLV-----GSV	982
QY	887	V--DDLVNPSFLDVLILSAKYEAANKITIRYKEKRNILKGREVYTD--KHQTTICAIRYS	942
Db	983	TKNDKASHSLDLADVLAAKTAALAEIKRYKQOONNI---OLTFPIQOKDEAMCAVRAYS	103
QY	943	FADIGLIIIGRDLRNRNDMYALQGHETFTVGNHKSJLNG-KGMDRYKND--APRYKLI	998
Db	1039	FADIGLIIIGRDMDEKSSJDMERLITTYKNIKKEHNDGIKDIPKTYGSESKPPAYKLI	109
QY	999	RENWNEANFAKYEAMKCDIYILADKSGHOSTOSSYGSQSDHTPLDYIPIOKIRMTTEMA	1051
Db	1099	RADWNEANHQQVYRAMKC-----ATKGITIC---PGMPVDYDIPIQLRMTTEMA	114
QY	1059	EWYCVOKKEVDKIKERKECKODNOGCGIKREGTGTCTEACNEYNDITOLMEQONN	1111
Db	1144	EWYCAQOSEDEKLIKKTADMSGDDG-CY-QGDVDCGCKAKACDKYKEIRIKWNEQWR	120
QY	1119	IISDYKKELEHQOASVNSNGSIEASTAKNIHD---BNVTEISETLYOONGK-----SMK	1177
Db	1202	KISDYKNLTYLQAKTSTNBP-----RYVLDDDDPDYQOAMDPLTPIHKASIAARULVKA	125
QY	1172	SSTSPESAVIGTNTTYENVGAYLHDTGNFDDCOSONEFCDEK-----SDGKDNEYAAR	1227
Db	1258	ASSPEIILAAAPITPPYSTAGYIHQETIGYCGEOFOOTFCEKIKHGASTSTYENKREYTER	131
QY	1226	DKPOHDHACCCGSKGSKPTBQVITKTKKKAERKDECTVNDILKENDGKOVYDCHPKN	1287
Db	1318	QPPPYATACDCINRSQYE-----EPKKENVESACKYVEKILLEGKNRGRTYGECPKKS	1377
QY	1286	SNGYTDMQC-GNINLVEDPRVCMPPRROKLCVHFPLANDNEIKKLOLKEAFTKSANA	1344
Db	1374	---YDMCKNNNDIISH-GACMPRRROKLCITYIAHESQTENIKTDNMLKDAFTKATA	1422
QY	1345	ETFEFMYWYKSK-DGEGNELDKELKEKIPAPLRFMSYTFGPDYRDLFTGDISGHGG	1407
Db	1430	ETFLSMOYKSKNDSEAKIIDR-----GLIPSQRLSRMATTFGDYRJOICLNTDISKONDY	1487
QY	1404	SKLRQJDSLFLKNGDQSPNGKTRQEWMTESHEIWEAMICALVLTGAKKDD---FTENY	1460
Db	1486	AKADKIGKFPKSKDQSPGSLSGQEWMTKNGEIKWGMICALITKYVTDNKRKRIKNDY	1546
QY	1461	GYNNVKSFSK-STLLEFAPKPOFLRMLTBMWYDYCYTQKYLAKDVQEKKS-NDOLKCD	1518
Db	1546	SYDVYNOSONGNPSLEEFAPKPOFLRMLTBMWGEFCAERQKKEKNTIKDACNEINSTOOCN	1605
QY	1519	T---ECKNKEDYKVKK-AKKEMITPDQKYKDEPKKRFEDROHIGVMTDVTGTNADY	1574
Db	1606	DAKRCNOACRAQYQYENKCKEFSQGTNNFVLKANVQYQODEYKGYEKD-----	1656
QY	1575	LNKRFTASCGDKQGSASVYONRLOLLEKQAYYADAKHGCCTYFIENDKYTNISSKDCK	1634
Db	1657	-----GVQF-----IGNNEYLQK---CDNNK-CSC-----	1678
QY	1635	GLVKEANTGAIKQWKNKGPNNYNNLKELTEDVLFPSRLICIFHALDGNVTPREYDENG	1694
Db	1679	-----	1691
QY	1695	RKRLMEVAALTEGYNLGOYTYKEKKEKIKTSDAHKYSYEVPPCSAMKSYFYDLRIDILGI	1754
Db	1692	-----PFGKY-----AHKYP-----	1701
QY	1755	DNLEDEKOTTEMLKTLFKNNKNGSVGKSDSTTGNGPSTARAKFFWNEKCEWMAAMIGY	1814
Db	1702	-----	1701
QY	1815	KRGHDGNSGNSARSDEDLKCGSVSPDDYDPMGNRDEGTAYOFLRWFAEMGEDFCBK	1874
Db	1702	-----	1701
QY	1875	EKELEKTVACNDYTOGDNEDKRRKCTDACTOYKKFISEMKPQYKEQOIKTYGENKDKIYS	1934

```

Db 1702 -----EKC----- 1704
QY 1935 EHVADAEAREYLDKOLKICENKSGCEYCMKDVSTQRLTDGNSQNPASLDEPK 1994
Db 1705 -----DC----- 1706
QY 1995 EYEGKCCQVPRGPRVRRRTPSPRVLISKATASKEAKTAPPTKQPKVENTTEMRA 2054
Db 1707 -YQK--HVSIPP----- 1725
QY 2055 QTRRAAQQTRKRTSTATTESDVGTMVKAISNKPDSRGLEGCPNPKTYGQ--PKWG 2112
Db 1726 -----PEAPVTVDVCSIVTKFK--DTNFPSPACGLK-YGKTAPESSWK 1766
QY 2113 CT-----VGSKENENGICMPRRKKLCINNIQYLNFT-----ENKRD 2151
Db 1767 CIPSDTKSGAATWTKGSGSGSICIPRRRRRLVYVKIQ--EWATALPQEGAPASHSRA 1824
QY 2152 NDIKEAFKCAIETQGLMKYIIE-NPAENE-----LQNGT 2188
Db 1825 DDJRNAFIQSAIETEFIMDRYAEKKPQDGSQOALSQLTSTYSDEEDPPDKLLQNGK 1884
QY 2189 IPDEFKIMTYTGYDYKDMFPGTDSNDKLIITVNSVTTLNENKKNKODKKDELRK 2248
Db 1885 IPPDFILMFTYTGDIYHIGNTSDSGNTSGNSNNNTVLEASGKEDMKQKQEKIEQ 1944
QY 2249 I-----FWEKNKFTWEGMIGLVY-----HUTDENEKIR 2280
Db 1945 ILPRNGSTPLVPKSAQTPDKMWEHAEISIMKICALTYTEKNPDTSGARDEKRIKDD 2004
QY 2281 DNY-----QYN-DMTKL-----TPSLEEFYKRP 2302
Db 2005 EYVEKFFGSGADHGTASTPTGYTKQYDEKVALDSTGAKTPSASSDPLLSDFVLRP 2064
QY 2303 QPLRMFTMAEFCNKRREOLIKLEAGCKEY-----CN---GSN 2339
Db 2065 PYRYLLEMGONCKKKHLLAQIKHECKYENGGSSRRGGITROYSGDEACEMLPKN 2124
QY 2340 DGTQOE-----CABACTYTNFKMKTEYEROREKFKKRD----- 2376
Db 2125 DGTVPDLEKSPKPCSSYKRWIESKGEFEKQKAYEQKQCVNGSNKHNDGFCETLT 2184
QY 2377 -GKKYDY-----PSTRDLEKATCAHEYLMLKELCGKDKSCMKPSSQLPKTQOS 2430
Db 2185 TSSNADEFLTLGPKCPNNVEGKTFDDDTFK-----HKKDDPCLKFSVCKK----- 2234
QY 2431 QSSDANDMPESLDYVPEEFKCEPSELKSGSMITHKTKTEPKIPMNCVEAAAYLSKEA 2490
Db 2235 -----DECD--NSKGTDCRNK-----NSID-----ATDI 2286
QY 2491 ENNMDTLKEKFTPIESTKKEKESKNSWTNNPCDPKRPVADKYIGRRNCPEN----- 2543
Db 2257 ENGVDSTV-----LEKRVASDSKSGFNGD-----GLENACRGAGIEGI 2295
QY 2544 REENRFVQVDEMYCKSKRYOEEKRYCVPRRHEMCLRLNDEIKIRLDSNVLAMVR 2603
Db 2296 RKD-----EMKC-RNVGCTV--VCKP----- 2313
QY 2604 RTARNEGIDILIKNSENENGCAMPICDTMKYSFADLGIYRGTDMLRIGLYPVEIKLY 2663
Db 2314 -----ENVNGE-----AKGRIIIGIRLNVK-----RMV 2336
QY 2664 KVEEYIYGKRNKNGKNGKND-----VQTRSAMWANDKRDIMKAMTCKAPKEDAKLFRKG 2719
Db 2337 EYEFEDNKKIKHISRIKNGEISPCIKNCVEKVVQDKRKE--WKEITERED----- 2387
QY 2720 RMDGFEITLIDKCGKDDPVDYIIPQFRWMTWSEYCYKALMELEFKKSCDHCK 2779
Db 2388 -----QYKNDSDDNVNSFLETILIPQITDAN--AKNKVTKLSFGNSCG-CS 2432
QY 2780 TSDCKNDYDNKCEQCKTRCOEYKKNFLVWKSLFDIOSNKKYELBQPIYTKLSTYDHW 2839
Db 2433 AS--ANEQKNK-----GEYKDAI-----DCMLKKLRD-----KIGCEBKK 2465

```

```

QY 2840 QNEVQKLTFRKSGCSVSFSEYLETSCNLNFKNENDGSSNIFTVAEETPKSYKE--- 2896
Db 2466 HH-----QTSPTCECSDPPQPTLED--ETLDDIETEAKKNMPPKICENLVKTAQODE 2518
QY 2897 -ACSTLPSKNPLDNCPTDONKD 2918
Db 2519 GGC--VPAENSEPAAIDSGKE 2538

RESULT 4
US-09-210-288-12
; Sequence 12, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnan, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xia-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210, 288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH21.1EMDVI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-09-210-288-12

Query Match 15.3%; Score 2972.5; DB 4; Length 2710;
Best Local Similarity 26.5%; Pred. No. 1.3e-203;
Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;

QY 9 IIKGNASSLEGDAPKSPITKESKRSANVLEARYAKNIRHPSK-YAKEHYDSLKGLDTKA 67
Db 6 LAKGPREAAGGDIED-----ESAKHMDRIGKDYDYKVEAEKRGGLGRLSEA 58
QY 68 EFRGSGSTPVNKHNYTTPYPCNLDKHEHTNLRIDVNLRHPCRGHREONRDEDESESC-G 126
Db 59 KFEKNESDPQTPED-----PCDDHKYHTNV--TTNVINPCADRSVRSDEYGGCCTH 110
QY 127 NKINRYKNDALICAPRRRHMCCKNLLEALNDINTQNIHDLNGLVYATAEGESTYNN 186

```

Db 111 NRIDSGOGDNKACAPRYRLHVCDQNLQJEPKIKITNTNHLVYVCMAMKFEQSITUD 170
 QY 187 HP-HKGT-----SDACTALARSFADIGDIVGIMDKERPNVHD-----KVENGLREVRKI 235
 Db 171 YPKYQATYGDSPQICITMLANSFADIGDIVGRRLIYGLNPEIOTROOLOENLNTITIGKI 230
 QY 236 HD-----GMEDEKKNYNDPGSGNYTKLRBAMVNNRNKWEAITCDASYSGYFMOBSNT 292
 Db 231 YEKINGAEARYGND--PE---FPKLEDMWTANRETYWKAITCNA--WGNTYF--HACNR 282
 QY 293 PLFSPNPKGKHOGKVPYTNLDVPOYLMPEDWEGEGRKRNKIKAKYKDSR-----NDKE 348
 Db 283 GERKKGCRDNDVPTYPYVPOYLMEWEMADFCRKNKIKDYKRCRCKDKDKD 342
 QY 349 RLYCSHNGDCTTITMKGILHLNDKCTDSTCKKVEVWLGNOQAEAFKREKEREKI 408
 Db 343 R-YCSRNGYDEKTKRAIGLRYGKCISCLYACNPYVDMINNOKEFDQKKYDEIK 401
 QY 409 SY-----LSMDNKFVNINSE--YKQPYELKETOYATNDPFLNLINEGRYC--- 454
 Db 402 KYENGASGSRQKRDAGGTTTNTDGEKKEFYELNKSERYVDKFEKLSNEEICTKVK 461
 QY 455 --KGLPEKEDITFTNSA-----DDKGFYRSEYQVCPDGVKODGKITYH--KSDND 504
 Db 462 DEEGTIDFKNVMSDSTSGASGTNVEGQFYFSKTCQPCPYCGVK---KVNNGSSNE 517
 QY 505 RERYNN-----EDYKPPMGVKNPTNITVLYSGNEQDGTOKLENFCN-----SS 547
 Db 518 WEKRNKCKCKSGKLYEKPDKEGTITILKSGKHDIIEELKPKCEKNGDITINSGGSG 577
 QY 548 TYNKDKNN-----OKMECYKROENINRCKLEONTEINNDKN-----IT----- 586
 Db 578 TGGSGGNSRQELYEKMC--YKGEDVYKVGHDDEDEDEENKNAAGCITLKNOKNKE 636
 QY 587 -----SFHNEFLMWTYLLDITKNDKIKTCI--NNTTTHC--IDEENRMC 629
 Db 637 EGGNTSEKEDETOKEFNFPEFYVAHMLKDSJHMKKILQRCLOQNRKLCGKNKKNNDNC 696
 QY 630 LCPDRVWOKKEEEMNSIKKLTFTKKNIQ-----QSYSTNNLEFGYF--- 672
 Db 697 ECKKRIOTOKDEMGIYOHF--KTONIKGRGSGDNTAELIPEDHDVLYQNLQEEFLKGD 755
 QY 673 -----FKYMDKLDKDEAKMKELMENIKRKNNEFSMLENNND-----YLENATELL 718
 Db 756 SEDASEKSENSLDAEA-----ELKHLREIIESEDNNQASVGGVTEOKNINDKLL 809
 QY 719 DILKERTATC-----KDNTEACETSHNATNPVCKPRGCTO--PTKNIKEIAO 766
 Db 810 NYEKDADLCLEIHEDEBEKEKGDNECIEEGENFRYNPCSGESGNKRYPIVANKVAYO 869
 QY 767 YFRKSAYEARNRGLHLKLGKAHEGYKRGGRKDEKMLCLMIKHSRNATGFSNGPCD 826
 Db 870 MHKKAQTOLASRAGSALSGLDISLOFKNGRSTLKGQICIKINENYSDNSKNGSGPCT 929
 QY 827 GKGTGDIOTRFEVGTWEVDEPHMKKDEEDYIMPRRHICTSMLEHJQTDHPLNGNI 886
 Db 930 GK-DGHHGVRRIRIGTEMNIGKQOTSYKNVFLPRRHKCTSMLENIDV-----GSV 982
 QY 887 V-DDLVNNSFLDVLISAKYTEANKITIRMKKKNKLGKREVTDP--KHQTTICRAIRYS 942
 Db 983 TKNDAASHSLDVOIAKTDAAELIKRYKDQNNI-----QLTDPYIOKQDEAMCAVAYS 1038
 QY 943 FADIDIDIRGRDLMERNGDVWLOGLHLETVENHISLKG--KGNDKYND---ARKYTL 998
 Db 1039 FADLDIIRGRMDWEDSSIDMETRLITVEFNKIEKHDKIDKNRYGDESCKRPAVYKL 1098
 QY 999 RENNWEANRAKWEAMKCDIKYLNKDSGHOSYOSYSGYSDHTPPDDYITPORKLMTMA 1058
 Db 1099 RADWMEANRHQVWRAMKC-----AKGIIIC---IGMPVDDYITPORKLMTMA 1143
 QY 1059 EMYCKVQAKKEDYKLEKECKEDKNDGOGCTRESGTGCTKTEACNEYNDIITGLMEQWN 1118
 Db 1144 EMYCKAOSQOEDYKLIKICADCMKSGDK--CT--QGDVDCGCKAKCADKYKEIEKMEQWR 1201

QY 1119 IISDKYELHEQOAMYSNGSIEASTAKNHD--RNVIEFLSELYOONGK-----SNK 1171
 Db 1202 KIDSKYMLIAQAKTTSTNG-----RTVLGDDDDPYQOWVDFLPIHKHSIARVLYKRA 1257
 QY 1172 SGTSDSAVIGTWTYENVAGYLHDGNDPDQOSQNEFDEK-----SDGADNEYAR 1225
 Db 1258 AGSPTEIAAAPTTPYSTAGYTHQETIGGCOEQOPEKHKGANSTSTTEKNEYTER 1317
 QY 1226 DKPDHDGACGCKSGSPTRVQITKTKKAEKDEKTEKYNDILKENDKROYEDCEPKKN 1285
 Db 1318 QPPEYATACDCINRSOTE-----EPKKKEENVSACKIYEKILEGNGGTGYGECPKKS 1373
 QY 1286 SNGYPMOC--GNINLYEDRVCMPPRRQOLCYHFLANDNEIKLOSQVNLKEAPITSA 1344
 Db 1374 ---YPMDCKNNDIDSHD--GACPPRRQOLCYIYAHNEOTENIKIDMLKAFITAA 1429
 QY 1345 EPEFSWYKYSK--DGBNELDKELKRGKIPPAFLRSMTFGDYRDLFGTDLISGHG 1403
 Db 1430 EPELSQWYKSKNDSAKTILDR-----GLIPSOFLSMATTFDYRDLCLNTDISKONDY 1485
 QY 1404 SKLEQIDSLFKNGDOKSPNGKTRQEMWTEHSHEIWEAMICALVYIGAKKD--FTEY 1460
 Db 1486 AKAKDIGKFFSKDGSKSPSGLSRQEMWTKNGPEIMKMLCALTYVYDTDKRKRIKNDY 1545
 QY 1461 GYNNVFE SDK--STTLEEFKAPQFLRLTEWYDDCYTRQKILKVOEKCS--NDOLKCD 1518
 Db 1546 STDKNVQSGNPSLEEFKAPQFLRLTEWYDDCYTRQKILKVOEKCS--NDOLKCD 1518
 QY 1519 T---EONKCEDYVYKMK--KKKEMIPODKYKDEBDKRFROHIGVWVDTGTGNADY 1574
 Db 1606 DAKHRCNQACRAYOEVENKKEKFGSGOTNNFVLKANNVQPODEYGYEYKD 1656
 QY 1575 LNRKFTASGCDRPGASAVYORNIOLLEKOAYDADKHGCTKFIENDKYNISSKDKR 1634
 Db 1657 -----GVOP-----IQGNEYLLQK--CDNNK--CSC----- 1678
 QY 1635 GLVKEANTGALIMONKGNPNYNNLKEITEDVLFPSRLRICEPHALDNGYTPREYDENG 1694
 Db 1679 ----- 1691
 QY 1695 RKLMEVATGESYINIGYKKEKKEKIKTSDAHKYSYEVPCSAKYSFYDLIDILGI 1754
 Db 1692 -----PRKY-----AHKYP----- 1701
 QY 1755 DNLDEKQTEENLKKIKNNKSTSVKSGSDSTGPNGSTAKFFMENKECYWAMATCGY 1814
 Db 1702 ----- 1701
 QY 1815 KRGDDNGSNGARSDEDLKRCGVSDDDYPMGNRDEGTAQOLFRLWFAEMGEPCPKH 1874
 Db 1702 ----- 1701
 QY 1875 EKELEKLVGAONDYTCGDNKRRKCTDCAQYKFLISEMPOYEOIKRYGENKDKIYS 1934
 Db 1702 -----EKC----- 1704
 QY 1935 EHPYAKDAEDAREYLDQKLIKICENKSGDCEYKCMKDYSTORLTDGSONMPASLDEPK 1994
 Db 1705 -----DC----- 1706
 QY 1995 EBEKSCNOVGRPPRVARETPSPRVLSISKAFSKKEAKTAPTKOPKVENLTEMRA 2054
 Db 1707 -YQK--HVSIP----- 1725
 QY 2055 QTRRRAAOQTRKRTSTATITTESDYGTMAVAILSNKPDSSNGGLEGNPKTYGYO--PKWG 2112
 Db 1726 -----PPAPVYTVVCSIVYTLF--DTNNFSDAGLK--YGTAPSPSK 1766
 QY 2113 CI-----YKSKENDNGICMPRRKRLCINNIOYLANET-----EKRD 2151
 Db 1767 CIPBDYTSAGATTKGKSGSDSGSICIPRRRLLYVGLQ--EWATALPOGEAGAPSHSRA 1824

QY 2152 NDIKEAFICAAIEPQPLKYLIE-NPAANE-----LONGT 2188
 Db 1825 DLRNFIAISAAIEEFLMDRYEKKPOGDSQALSLQSTSYSDDEEDPPRLKLLONG 1884
 QY 2189 IPDEFRIYVYGDYKDMFEGDISNDKLIIVTNSVTLLENKKKODKDELRK 2248
 Db 1885 IPDEFRIYVYGDYKDMFEGDISNDKLIIVTNSVTLLENKKKODKDELRK 1944
 QY 2249 I-----FWEKNKFIWEGMIGLYT-----HITDENEKIR 2280
 Db 1945 ILFKNGSTPLVPRKSAQTPDKMNNHAEISWKMICALYTEKNPDTLSANGDENKLEKD 2004
 QY 2281 DMT-----QYN-DMTKL-----TPSLDEEYKRP 2302
 Db 2005 EYKEKEFGSTADKNGTASPTGYTKQYDEKVKLEDTSGAKTPSASDPLLSDFYLRP 2064
 QY 2303 QFLRMTEMAEFCNRKBDLLEAGCKEY-----CN---GSN 2339
 Db 2065 PYRYLEMGONCKRKRAKLAQIKHECKVEENGSGSRGCTROYSGDEACNEMLPKN 2124
 QY 2340 DGTQOE---CAEACVYQNFILKKWTEYERQERFKKDKD----- 2376
 Db 2125 DGTVPDLKSPKAPCSSYRKWIESKGEFEKEKAYEQCKRCVNGSNKHDNGFCETLT 2184
 QY 2377 -GKKYDY---PSTERDIEKATCAHEYINMLKELCGKDCSCMQKPSQPLKTTQOS 2430
 Db 2185 TSSKADFLTKTLGPRCPNNVEGTIPDDDKTFK---HTKDDPCILKFSVNCCK--- 2234
 QY 2431 QSSDANDMPESLDYVEEFCPCPELSKSGSIHTKRIEPIPMNCVKAAYISKEA 2490
 Db 2235 -----DECD---NSKGTICRKN---NSID---ATD1 2256
 QY 2491 ENNMDTLKEKFIPIESTKESKSNMTNNPCDPKRPADKYIGRNPCEN----- 2543
 Db 2257 ENGVDSTV-----LEWRVADSCKSGFNGD-----GLENACRGAGIEPI 2295
 QY 2544 REENRKYVDEEMKCYKSKYQEKKRCVPRREHMCRLNDELKIERLSDNYLAKMR 2603
 Db 2296 RKO-----EMWC-RNVCYV---VCKP----- 2313
 QY 2604 RTARNEGIDILKFNSENGCAMPICDTMYKSPADLGIYRGTDMLRIGGLPVEIKLY 2653
 Db 2314 -----ENVNGE-----AKGHIILQIALVR---RWV 2336
 QY 2664 KVEEYLYGKRNKRNKRNKYND---VQTFRSAMDMANRDKWAMTCKAPEDAKLERKG 2719
 Db 2337 EYEFEDYKRIKIRIKIRIKGEISPCIKNCVEMKQKRE-KWEITERFKD----- 2387
 QY 2720 RMDGFERITLIDKCGHNDPRVDYIPOFRMTEMSEYCCALMELEKFKSCDHCK 2779
 Db 2388 -----QYKNDNSDDNVSFLETLIPQITDAN--AKNKYIKLSKFGNSCG-CS 2432
 QY 2780 TSDRCNDYDENKCEOKTCEQYKNFVLWKSLFDIOSKRYEIQPIYTRISTDYH 2839
 Db 2433 AS---ANEQKN-----GEYKDAI-----DOMLKLLD-----KIGBEKK 2465
 QY 2840 QNFVQKLTFKSCSVSESEYIHEYSKCLNYFNENOGSSNIRYAFEEETPKSYKE--- 2896
 Db 2466 HH-----QTSDECTSTPOQOTLED--ETLDDIETEAKNNMPKICENVLTAQODE 2518
 QY 2897 -ACSTLPSKRNPLNDCPTQONKD 2918
 Db 2519 GGC---VPAENSEEPATDSGRE 2538

RESULT 5
 US-08-487-826B-16
 ; Sequence 16. Application US/08487826B
 ; Patent No. 593827
 ; GENERAL INFORMATION:
 ; APPLICANT: Slim, Kim L.
 ; APPLICANT: Chitnis, Chetan
 ; APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMIDIUM VIIVAX
 TITLE OF INVENTION: AND PLASMIDIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487,826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121,001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ. ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2182 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal
 US-08-487-826B-16

Query Match 9.08; Score 1743; DB 2; Length 2182;
 Best Local Similarity 20.38; Pred. No. 1,4e-115;
 Matches 757; Conservative 339; Mismatches 678; Indels 1960; Gaps 126;

QY 13 GNAASSLEGDAKSPILKESHKS-ARNVLER-----YAKNIRHPSKAKEHVSLSKGLTK 66
 Db 10 GSGGSS-SGCKGKDDSYIIVSDAKDLIDVGEKVEYKRNKD--AKKTIKALKNLNT 66
 QY 67 AEFRGCPSTVYKHNYYYPYPCNLDRKHTNLRDYDVN---LRRPCH---GREQNRFE 119
 Db 67 ANGRSSEFASIE-----TCTL-VKEY-----YERVNGDGKRRHPCRKDAKNEDVNRFS 114
 QY 120 DESEEGC-NKIRNYKRRNDAT-ACAPRRRRHMDKNLEANDINTQIHHLLGNVLTAK 177
 Db 115 TLGGQCTYNRKIDSOQDNVACAPYRLRLHLDYINLESID--TTSYTHKLLLEVCAAK 172
 QY 178 YEGESIYNNH--PHKGT-----SDACTALARSFADIGIYRGIDMF-----RPNVADKY 224
 Db 173 YEGNSTI-NHYTHQHRKNNESASQLCTYLARSFADIGIYRGIDYIYGYDNKEKEQKKL 231
 QY 225 ETLGLEYFKKIHDMGDEVDKND-----YNPDG--SGNYIKLRAMWNNVNRKWEALITGDA 278
 Db 232 EOKLKDELFKKIH--KDVMTNGAQERYIDAKGGDEFQLREDDMTWSNRETYWALICHA 288
 QY 279 SYKGYFMOSSESNPLFSNKGCHGKQKVPNTINDYVPOYLRWFDENGESECRNRNKKK 338
 Db 289 PKRANFIKTACVNGKCTNQCCHCIGSDVPTFYDYVPOYLRWFEWADECKRRKKKLEN 348
 QY 339 VKDSCRNDKERLYCSHNGHDCTTTIWKKGILHLDNKCTDCSTCKYFVWVGNQDEAFKR 398
 Db 349 LQKQCRDYEQNLVCSGNGYCTKTIYKKGLVIGEHTGNSVNCRMETWIDNQKKEFLK 408

```

OY 399 OKETKEI-----QSLSDNKNFNINSEYKOFETKEFOYATN 441
Db 409 QKRKETEISGGSGSKSPKTRKARARSSSDN-----GESKRYKRLKEVGOV 460
OY 442 DFLNLMGKYGK-----GLPGEADITFTNSADKGLFTYSEYQVCPDCKGDIKY 497
Db 461 DKFLKILKEGICQOPQOVGNEKADNVDFTNEKYK--TFSTRTETCEPCWGLKKEGPGW 519
OY 498 THKSDNDRERNEDYKPPMGVPTNITVLVSGNEOGDITQKLNPNSSSTNYDKNNK 557
Db 520 KVKGDKTGSAKTYDPK---NITDIPVLKPDSSQONILKRYNFEKGP--GGGQIKK 575
OY 558 MECYKID-----ENINCKLEQNTENINDNKKIISPHNFELWYLLRDTIKWMDIKT 612
Db 576 WQCYDEHRPSSKNNNKNCVEGTWDFTOGKQYKSYNVEFWDMVHMLHDSVEKTELK 635
OY 613 CINTTT---HCIDECNRNCICEDRWVKOKEEEMNSIKLFTKKNIQOSYSNINNF 668
Db 636 CINNNTGNCRNNKKTGCGQKQVKEKKQDBMAIKDHFQKQDIDVQ---KGLI 690
OY 669 EGIFKYMKLDKDEAFKELMENIKRKNFESNLNNRDLT-----NAI 714
Db 691 VFSEYGLDLY---LKGMLNLIKDYHGDITDIIKHIIKLDEDAVAVVGLGKXNTTI 746
OY 715 ELLDHLKETATICKDNNTNEACETSHNATNRCVPRGTOPTKNIKEIAQYKRSAYE 774
Db 747 DKLLQHKEDAEQCKK--OECE-----KKAQ 773
OY 775 EARNRGLHKLKRAHEGIYKRGRRKDFKDLNCRIMIKHSNMLGFSNGPDKGTGDI 834
Db 774 ESRGRSAE-----TRIDER-----TQOPDASGE----- 797
OY 835 QTFVVTGEWEDVPEHMKRKHEDVIMPPRRRICTSNLEHLQTDHPLNGNIVDDLNN 894
Db 798 -----VEEEDDDYDEDED-----DVOQOE 820
OY 895 FLGDVLISAKYEANKIIRMYKEKNLKGPEVTPKHOITICAIRYSPADIGDIIRGD 954
Db 821 EEG-----KEBGIVTEVTEVEVEET----- 842
OY 955 LWERNDMVKLOGLHETVFNHLSLKGKNDKYNDAPRYKLRENWMEANRAKWEAM 1014
Db 843 ----- 842
OY 1015 KCDIKYKLSGHQSSQSSYCGSDHTPLDDYIPQKLRMTEAEYCAVQKKEDKLE 1074
Db 843 -----VTE-----OE 847
OY 1075 KCECKDKNGOGCTKESGTGCTKEACNEYNIIIGLKEQOMIIISDKYKELHQAQMS 1134
Db 848 GVKPC-----DIVGKL----- 858
OY 1135 VANSGEASTAKNHTDRNIVIEFLSELVYOONGSKNSGSTDESAGVIGTNTYENVAYL 1194
Db 859 ----- 858
OY 1195 HDTGNFDDQOSQNEFCDEKSDGDKNEKYAFBDKPDHDGACGCKSGSKPTRYQITKKKA 1254
Db 859 -----ACGLKTYP----- 875
OY 1255 EEKOTECTVNDILKENGGKOVEDCHPRKNSNGYPRMOCGNINLV-----EDPRVCM 1308
Db 876 -----GKER-----FPNMKCVTPSGSVSTATSGKQALICVP 906
OY 1309 PRROKLCYHFLA---NDNEIKRLQSOV-----NIKEAFIKAAAEFTFSWYTY 1353
Db 907 PRRRLLVYGGISQWASRGDDETTVESSSEATAPSOSESEKRLTAFTIESAALETFFLMHY 966
OY 1354 K-----SKDG-----EGNELDKELKE--GRIPAPAFRSMYFTGDRDPLFG 1393
Db 967 KEKKPRATQAGAGLGVSLPEPSPPGEDPQJQOQITGVTDFLRQMTYTLADYKDLI-- 1024
OY 1394 TDISGHEGSKLKEQJLISLFRKNGQSPNGKTRQEWMTESHSHEIWMALCALVIGAKK 1453

```

```

Db 1025 ----- 1024
OY 1454 DDEFENYNNVAFSDKSTLFEFAKRPQFLRMJLTEMWYDYCTROKYLKDQEKCKSND 1513
Db 1025 ----- 1024
OY 1514 QLKCDTBCNKCEDVYKWKKKKEWIPQDKYYKDERBKRPRDROHIGVMTDTGTNAD 1573
Db 1025 -----YSSNDIS 1032
OY 1574 YLNRKFTASGDKPGSASVQNRNIQLEKQAYYDADKHCCGCTKFIENDDKYNNISKDC 1633
Db 1033 -----DYTGKQPSSSNML 1047
OY 1634 KGLVKEANTGAIKWNKGNNYNNIKELTEDVLFPSRLRICFHALDGNVTPPEVKDENG 1693
Db 1048 KNIVLEAS----- 1055
OY 1694 LKRRLMEVATEGYNLGOYKKEKKEKIKTSDAKRYTEVPPCSAMKYSFYDLRDIILG 1753
Db 1056 -----GSTE-----OEKEMK----- 1066
OY 1754 IDNLEDEKQTEENLKIFNKNGTSVGKGSQDSTGNPGSTARKFPYNNKCEYNNAMTCG 1813
Db 1067 -----QIQAKIKILIN--GATS---GVPPYTNASYKTPQOITWENIADIMWAMVCA 1113
OY 1814 --YKGRDDGNSGNSARSDDELKCGVSPDDDYPMGRNDEGTAYOFLRWFMEWGEDFC 1871
Db 1114 LTYKE-----NDAR-----GTS----- 1125
OY 1872 KHKRELEKLVGACNDYTCGDNEDKRRKCTDACTQYKFFISEMKPQYKQIKKYGENDK 1931
Db 1126 -----AKTEOUND- 1133
OY 1932 IYSEHPVAKDAEDAREYLDKOLKICENKSGDCYKMKQVSTORLJTDGSONNPASLDD 1991
Db 1134 -----LKK-----ALWDEANKNP----- 1147
OY 1992 EPKEVEGKNGQVPRGPRVAVRETPTSPRVSLSIKATSKKEAKTAPPTKPKVENLTTE 2051
Db 1148 ----- 1147
OY 2052 MRAQTRRAAQORRTKRTSTATTESDVTGWKAILSNKPSRSGICGCPKTYGQYPKW 2111
Db 1148 ----- 1147
OY 2112 GCITYGSKENENGICMPRRKLCINNIOYLYNTEENKRDNDIKAFITCAAIETOPLM 2171
Db 1148 -----IERKYQYNNVLE----- 1159
OY 2172 KYIENPAENELONGTIPDEFKRIIMYTYGVDYDMFEGTDISNDKIIITYTSVTIILN 2231
Db 1160 -----D 1160
OY 2232 ENMKKODKKDEBLRKIEWENKKNFIWEGMITYGLTYHLJDNENEKIRIDNYQINDMTKL 2291
Db 1161 ESGAKSND-----TID 1171
OY 2292 TPSLEEVKRRPOFLMPTTEWAEERONKRLKLEAGC-----KEYCNGSNDK--- 2342
Db 1172 PPTLKNFVEIPTFRMLJHEGNSFCERARLQAIHCEDEDEGEKQY---SGDEYCE 1227
OY 2343 -----TOCEACVYQNEFIKMTETEYEROREKFKDKDGKYYOYPTSTED 2389
Db 1228 EIFSQYVNLQDSSQAKRCRLKTIWIEKKTYEKQKAYHQQK---SVTEHQRD 1282
OY 2390 IEKATCAHEYLANKLELGCNKKDCSCMQKRPSSQLPPTTQOSOSDANDMPESLDYPP--E 2447
Db 1283 -----KC-----QOSNNANNESSRLGASPTA 1306
OY 2448 EFNKCEPBLSKKSMIHTIKITEPIPMNVCAKAYIYSKEENNMMDITLKEKPIPIES 2507

```



```

Db 1307 EF-----LQKLS-----CKNDNGT-----ENGDNKIDK-NPDKT 1337
QY 2508 TKESKNSWTNNPCDPKRPYADKYIGRRNFCENRENFYDYEWMCKYKNSKFYQEK 2567
Db 1338 FKEAHS-----CDP-CPTGVK-----CQNGH----- 1358
QY 2568 KRYCVPRRRHMLRLNDEKIRLKDSTNLYLKMVRTANEGIDITKNSENGCAMP 2627
Db 1359 ---CVSANGCKEKNM--KTADIKN-----KTDNGNIEWVSDSTN----- 1398
QY 2628 ICDTKKSPFADIVGTGDMRLIGLYLPEVEIKLYKFEYIGKMRNKKGRKKYDVO 2667
Db 1399 ---TFPHLD-----CKSSGIFGI----- 1415
QY 2688 TFRSAMWANDNRKDIWK-AMTKAPEDAKLERGRMDGFERITLIOD-KCGHKDDPVVDY 2745
Db 1416 ---RKDEWKCANVC-----GVDICTLEKKIKNOGEGD---KRY 1447
QY 2746 IPRFRMWTWSEYCYKALMBELERKKSODHC-KTSDCKNDYDENCKCQCTRCQEK 2804
Db 1448 ITWK-ELTKRWLEYF---LEDYNRIKKIKLCTKCKEDGCK-----CJGCTE-- 1490
QY 2805 NFYLKMSLFEDISNKKYKEIYEPITKISTYHQNQFQKLTPEKSECVSESEYLHE 2864
Db 1491 ---KM-----VO---BTKEMO----- 1501
QY 2865 TSCKLAKFNENDSNIRTYAEETPKSYKACSCITLPSKNPLDNCPTQNDGCKELQ 2924
Db 1502 ---KIND-----TY-----LE 1509
QY 2925 TFFPCKNDYDNNILDMNNAVLYNSDDNKGLIPRRRHLCITRPITAYVRKGDKEILK 2984
Db 1510 QY---KNDGDNLTJTF---LEOFYR----- 1529
QY 2985 KLLTSAFSGQLGQYKKEEELCFEAMKSYADSDITKGTMDMTLSSEIKIKIET 3044
Db 1530 ---TFKNAIKPCDGLD----- 1543
QY 3045 SNEATENRKTWENNRRQIWHAMLCGYKATSKYTLDEGWCQLPKDEITNOFLMLIEMA 3104
Db 1544 ---QF----- 1545
QY 3105 KQACKKKHVSLSLTKCPKPSNEDNEFEASBLQPCQCONDIRKYSILNLIKMTMELNT 3164
Db 1546 ---KTSQGNSTDN---SONGNNDN---VLCILNKLQK 1575
QY 3165 KYKQLDSSGNIDNKPSSENVQSYKSDOCALENDINETYGKNENN----- 3217
Db 1576 KISECKEBOHSGQYQF-PCD---NSSLSGKSTLVEDDYDE---QNEPNKVEQPKFC 1626
QY 3218 -EFKEVYLKLLPYGLYVEDETHRNHVLGNI--KEEQVPRKALYFTPHVDSFYQAPL 3274
Db 1627 PDKPEPKKENDDEVGTCGGEKKYEDSVIEQKEEELAAAPESPPLTF-----EAPK 1680
QY 3275 FSTHRAQYDP-----KN-----DLKSSISVIVSALGL-----IALHEKKEKFS 3316
Db 1681 KEENVVPPKPPPKRRIRKIRNVLDPHAPVAPALMSSTIMMSIGGFAFTFYLLKTKTKS 1740
QY 3317 SV-DLRILNITPOGEYMGPTLESKNRPIYRSGPYGKGTIYMEGSDGDEYMDLSS 3375
Db 1741 SVNLKQIIDLIPASDIDIPLKSSNRITIPASDRHKGKITYIMEGSSGDE-KTAFMSDT 1799
QY 3376 SDITSSSEYEELDINDIYVPGSPKTYLLEVLEPSKRDIPSD-----DTPSNTPR 3428
Db 1800 TDTTSSSEYEELDINDIYVPGSPKTYLLEVLEPSKRDITONDINDIDISDIPNSTPP 1859
QY 3429 TNPFDIDEMELKADHYQYL---PNTPENNNTKASADIPMTEPNUTLYSDNPEKPPITIS 3485
Db 1860 P--ITDEMQLKRDFTSNMLQNTQNTPEPILHDNDV--NNTHP-TMSRNHMQKPFIMS 1914
QY 3486 IHRDLYTKEISY-----NINMSTN-----TNDIPNANANDSGRIGDILNDS 3529
Db 1915 IHRNLTSGEELIYDMFNSGNFIPNISDSTNSMDSLTNSHSPYNDKNDLYSGDILNDA 1974

```

```

QY 3530 LVVNLNLIY--MMK 3541
Db 1975 LSGNHIDIDYDMLK 1988

RESULT 6
US-08-568-459A-10
; Sequence 10, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xia-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-Dec-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelson, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-568-459A-10

Query Match 5.3%; Score 1032.5; DB 2; Length 700;
Best Local Similarity 34.8%; Pred. No. 2e-65;
Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDATACAPRRRRHMCNLEALNDINTONTIHDLGNLVYAKYGESIVNNH--PHKGT- 192
Db 5 NKVGACAPRYRLHLCDNLNLSID--TTSSTHKLLEVCMAKAYGNSI-WHYHQHRTN 61
QY 193 ---SDACTALARSFADIGIVRGIDMF-----KRVVHDKVETGLREVRKTHDGMDE 242
Db 62 EDSASQICTVLANFAIDIGIVRGKIDLYLGYNKKEQRKRLKDLDFIKTH--KDV 118
QY 243 VKND-----YNPDG-SGNYKYLREAWNVNKNKWEAITCDASYSGYFMQSESNTPLFS 296
Db 119 MKTGAQERYTIDAKGKGDFQLRDMWTSNRETYWKAALICHAPREANYFIKTACNVGKT 178
QY 297 NPKCGHQCQVPPNLDVYQYLRWFDEWGEFCKKRIKILAKKVAQSDGRNDRKRLYCSHG 356

```

```

Db 179 NGQCHIGDVPYTPDYPPQYLKMFEEAEDPCRRKKKLENLQKCHDYBQNLGCSGNG 238
QY 357 HDCTTTKKKGLIHLNDCTDCTCKAYEYEWLGNQDAFKEKREYKEI-----407
Db 239 YDCTTIYKKGLVIGEHCTNCSWCMRYETWIDNQKKEFLKQKRYETETISGGSGSKSP 298
QY 408 -----QSYLSDNKRNVNINSEYKYQFEKLETOYATNDFLNLNGKCYCKG---456
Db 299 KRTKRAARSSSSSDN-----GYESKFYKKLEVOYQDVDFELKILNKEGICOKOPQ 350
QY 457 -GLPGEKDITFTNSADKGIYFSEYCOVPCDGVCKGIGITYTHKSDNDRVRANNEDYKP 515
Db 351 VGNKEADNVDPFTNEKYK-TFSRFEICEPCWGLEKGGPWKYGKGTGCSAKTKYDP 409
QY 516 PWGKAPRITVLYSGNEGQDITOKLEFNCNSSTNYKDKNNQKWCYKYD-----ENINRC 570
Db 410 K---NIDIPVLYPKDSQONILKKYKNCCEKGA-PGGQJIKKQCYIDEHRPSSKNNNC 465
QY 571 KLEQNTENNDNPKIISFHNFEFLWYLLRDTIKMNDKLTCTINNTT-----HCIDECN 626
Db 466 VEGTWDKFTQOKQVKSINVFEDWVMDMLHDSYEMKTELSCINNTNGTCRNNKCK 525
QY 627 RNCICFPRWYKQKEEENSISIKLETKKNTQOQSYYSNINLFGYFPKVDKLDKDEAKY 686
Db 526 TDCGCFQKWEKKQOEWMAIKDHGKQTDLYOQ---KGLIYFSPYGVLDV---LKG 576
QY 687 KEMENIKRKKNEFSNLENNRDYLE-----NAIELLDHLKETATICKDNN 732
Db 577 GNLLQNKIDVHGTDIDKIKHLKLEDEDAVAVVYLGKDNNTTIDKLOHEKEQOEGQCK 635
QY 733 TNEACE-----TSHNATTNPGVPRGTOPTKNIKEIAQYKRSAYEE 775
Db 636 -QECECKKAQOESGRSAETREDEKQ---QPADSAGEVEEEDDDYDE 681

RESULT 7
US-08-487-826B-10
; Sequence 10, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Welliams, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VITAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelisen, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH21,001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-08-487-826B-10

Query Match 5.3%; Score 1032.5; DB 2; Length 700;
Best local Similarity 34.8%; Pred. No. 2e-65;
Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDALACAPRRRRHMDCKNLALNDINTQNIHDLGNVLTAYEGESYVNNH--PHKGT-192
Db 5 NKVACAPYRRLHLCDYNLESID--TSTTHKLLLEVCAAKYEGNSTI-NHYTOHRTN 61
QY 193 ----SDACTALARSFADIGDIYRGIDM-----KPNVHDKVETGLREYFKIHGMDE 242
Db 62 EDSASQLCTVYLRSPADIGDIDYRGKDLVLYGNKKEQKRLKLEOKIDIFKIH---KDV 118
QY 243 VAND-----YNPDG--SGNYKLEAMNVRNKRKVEATITCDASYSGVFMGSESTPLFS 296
Db 119 KMTNGAQERYIDDAKGDFQFQLEDWMTSNRETWKALICAPKAFANFTITACVKGKT 178
QY 297 NEKCGHKQGVNTNDYPPQYLRFWDENGEEFCRRKNIKLVKDYSCNDRKERYLCSHNG 356
Db 179 NGQCHIGDVPYTPDYPPQYLKMFEEAEDPCRRKKKLENLQKCHDYBQNLGCSGNG 238
QY 357 HDCTTTKKKGLIHLNDCTDCTCKAYEYEWLGNQDAFKEKREYKEI-----407
Db 239 YDCTTIYKKGLVIGEHCTNCSWCMRYETWIDNQKKEFLKQKRYETETISGGSGSKSP 298
QY 408 -----QSYLSDNKRNVNINSEYKYQFEKLETOYATNDFLNLNGKCYCKG---456
Db 299 KRTKRAARSSSSSDN-----GYESKFYKKLEVOYQDVDFELKILNKEGICOKOPQ 350
QY 457 -GLPGEKDITFTNSADKGIYFSEYCOVPCDGVCKGIGITYTHKSDNDRVRANNEDYKP 515
Db 351 VGNKEADNVDPFTNEKYK-TFSRFEICEPCWGLEKGGPWKYGKGTGCSAKTKYDP 409
QY 516 PWGKAPRITVLYSGNEGQDITOKLEFNCNSSTNYKDKNNQKWCYKYD-----ENINRC 570
Db 410 K---NIDIPVLYPKDSQONILKKYKNCCEKGA-PGGQJIKKQCYIDEHRPSSKNNNC 465
QY 571 KLEQNTENNDNPKIISFHNFEFLWYLLRDTIKMNDKLTCTINNTT-----HCIDECN 626
Db 466 VEGTWDKFTQOKQVKSINVFEDWVMDMLHDSYEMKTELSCINNTNGTCRNNKCK 525
QY 627 RNCICFPRWYKQKEEENSISIKLETKKNTQOQSYYSNINLFGYFPKVDKLDKDEAKY 686
Db 526 TDCGCFQKWEKKQOEWMAIKDHGKQTDLYOQ---KGLIYFSPYGVLDV---LKG 576
QY 687 KEMENIKRKKNEFSNLENNRDYLE-----NAIELLDHLKETATICKDNN 732
Db 577 GNLLQNKIDVHGTDIDKIKHLKLEDEDAVAVVYLGKDNNTTIDKLOHEKEQOEGQCK 635
QY 733 TNEACE-----TSHNATTNPGVPRGTOPTKNIKEIAQYKRSAYEE 775
Db 636 -QECECKKAQOESGRSAETREDEKQ---QPADSAGEVEEEDDDYDE 681

RESULT 8
US-09-210-288-10
; Sequence 10, Application US/09210288
; Patent No. 6392026
; GENERAL INFORMATION:
; APPLICANT: Slim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.

```


QY 1419 QKSPNG-KTROEWMTESHSHEWEAMLCAL-----YKIGAKKDDPTENYGYNNYKF-S 1468
DB 547 YKISNSIRYKRSWMTNGPIYMEGMLCALSYDTSLNANVPETHKLLTEGNNNEKYIFGS 606
QY 1469 DKSTTLEEFAPKROFLWMLTEWYDYCYTRQKYLKDVOECKSNDOLKCTECCNK---C 1525
DB 607 DSSTTTSKFSERQFLWMLTEWGENEFCCKEYKYLAKCKOD-VDGSGCKNGKCVAC 665
QY 1526 EDYVK-----YMKKKKEWIPQDK--YKDERDKKRPD--RQHIGVWVTDYTG 1569
DB 666 KDCKQYHSMIGIWMIDYKQKGRYTEVKKIPLYKEDKDVNSDADYLTQLONNKCV 725
QY 1570 NATDYLN-----RKFTASGDKRPGSASVYQRNIOLEKQAYYDADKHCCTKFEIENDK 1623
DB 726 NGTDECEYKCMHKTSTSDMPES-----LDERPEYKDK-CNC---VPNECN 771
QY 1624 YTNISS-----KDKCKGL 1636
DB 772 ALSVSGSGFPDGAFGGVLGTCGL 798

SULT 11
US-09-210-288-8
; Sequence 8, Application US/09210288
; Patent No. 6393026
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210, 288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-09-210-288-8

Query Match 3.3%; Score 633; DB 4; Length 921;
Best Local Similarity 27.8%; Pred. No. 1,4e-36;
Matches 224; Conservative 113; Mismatches 264; Indels 206; Gaps 42;

QY 948 DIIRGDLERNQDMVQLQCHLETVEGNIHSLKCKGNDRYN-----DAPKYLKLENN 1002
DB 80 DYLTASDI-ENGNSI-----GNIDMVVSDKANGFNGLDAGSANIFGIRKEQ 128
QY 1003 WEANRAVWAMACDIYLLDKSGHSTQSSYCYSHPTFLDYIPKLLMNTMAMAYC 1062
DB 129 WKC-AKV-----CGLDVCGIKNGN-----GSIDKQKIIIIAL--LKWVEYFL 171
QY 1063 KVQKREYDLKCKECKEDKNDNGGCTKESGCTGCTKTEACNEYNIDITGLMKEQNTISD 1122
DB 172 -----EDYNNKINAKISHTCKKDNSTCTINDCPNCT-CVE-----EWINKREIMAKIK 220
QY 1123 KYKELHE-----QAQMSVSN-----SGIEA-----SSTAKNH 1149
DB 221 HYQTQNGNDNNKSLVTIILGALOPSDVNRKAIKPSGLTAFESFGGLNADNSEKKEG 280
QY 1150 ID-----RNYEFSELYQONGSKNSGTSDESAVIGTNTTYNNGAYLHDTGNFD 1201
DB 281 EDYDLVLCMLNLEKQIQECKKKHGETSVENG--GKSCPTLDNTTLEE-----E 327
QY 1202 DCOSONE-----FCDEKS--DGKDNK-----YAPDRKPDHDGACGCKSGSKPTRVQ 1247
DB 328 PIBEENQVEAPNICQKQVEDKCKKEEBEBCCTPASPYPEKPYRH--YARRRTFTPEYVK 385
QY 1248 IKTAKKAEEKDTECKTVNDILKENDGRKQVEDCHPKKNSNGYPDQCGNINL-VEDPRVC 1306
DB 386 I-----WRGRNRKTTCEIYVAEMLKDKNGRTVGEYRKT-----YSEWTCDESKIKMGHGAC 439
QY 1307 MPRRQKLCYHFLANDIEIKKLSQVN-LKAPLTKSAAEFTFSWY-YYSKQDEGELD 1364
DB 440 IPRRQKLCYHFLANDIEIKKLSQVN-LKAPLTKSAAEFTFSWY-YYSKQDEGELD 493
QY 1365 KELKEGKIPPAFLSMYTFEGDYRDLFEGDI-----SKHGEGSKLKQDISLFGND 1418
DB 494 EKLKGILPEDEFRQMTYTTADYRIDIGLIDISKKTSGCVG---KVCNIDVF----- 546
QY 1419 QKSPNG-KTROEWMTESHSHEWEAMLCAL-----YKIGAKKDDPTENYGYNNYKF-S 1468
DB 547 YKISNSIRYKRSWMTNGPIYMEGMLCALSYDTSLNANVPETHKLLTEGNNNEKYIFGS 606
QY 1469 DKSTTLEEFAPKROFLWMLTEWYDYCYTRQKYLKDVOECKSNDOLKCTECCNK---C 1525
DB 607 DSSTTTSKFSERQFLWMLTEWGENEFCCKEYKYLAKCKOD-VDGSGCKNGKCVAC 665
QY 1526 EDYVK-----YMKKKKEWIPQDK--YKDERDKKRPD--RQHIGVWVTDYTG 1569
DB 666 KDCKQYHSMIGIWMIDYKQKGRYTEVKKIPLYKEDKDVNSDADYLTQLONNKCV 725
QY 1570 NATDYLN-----RKFTASGDKRPGSASVYQRNIOLEKQAYYDADKHCCTKFEIENDK 1623
DB 726 NGTDECEYKCMHKTSTSDMPES-----LDERPEYKDK-CNC---VPNECN 771
QY 1624 YTNISS-----KDKCKGL 1636
DB 772 ALSVSGSGFPDGAFGGVLGTCGL 798

RESULT 12
US-08-568-459A-4
; Sequence 4, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:

ADDRESS: Knobb Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/568,459A
 FILING DATE: 07-DEC-1995
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-0176
 TELEFAX: (619) 235-8550
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1435 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 US-08-568-459A-4

Query Match 3.2%; Score 617; DB 2; Length 1435;
 Best Local Similarity 21.8%; Pred. No. 3.9e-35;
 Matches 332; Conservative 215; Mismatches 557; Indels 418; Gaps 69;

QY 1101 EACNEYNIDIGLAKED--WNIISDKYKELHQAQMSVNSGIEASSTAKNHIIDRVYIEF 1157
 DB 20 KARNER-DI---KENEKELDYKKEKFNELDKKRYGNVQKT-----DKKIFEF 62
 QY 1158 I-SELYOONGSKNS---GTSDESAYVIGNTTYENVAIYLDHGNFDDCOSQNEFCE 1212
 DB 63 IINKLIDILNKFNRKMSYGTDPN---IDKMSLSIN----- 96
 QY 1213 KSDGDNKRYARDRPODHGACGSGSKSPRYVO-ITKTKRAEKEDIECKTVNDILKEN 1271
 DB 97 ---KHNEEMFNMMYQSPFSLTSSLIKQNKYVINAIVRSRLSLFDSR---INNGRNTS 149
 QY 1272 DGKQVDECHPRKKNNGYVDMQCGNINLVEDRYVCMPPRRQKLYCHFLANDNEIKKLSQ 1331
 DB 150 SNNEVLNMCNREKRG--MKMDCKRN--DRSNVYCIPIPRRIOLCLVINS---IKITYK 201
 QY 1332 VILKEAFIKSAAEFFSWYKSKDGEENELDKELGKIPAPFLRSMFYEGYRQFL 1391
 DB 202 EIMKHIFIEASKKE---SOLLKAKNDKNYS-----KCNCKINSFLDIYGHIA 246
 QY 1392 FGTDISKHGEGSKLEQIDSLFKNGDQSPNGKT---ROEMWTEHSEHTEWAMLCALVK 1448
 DB 247 MGNDDPG-GYSTKAENKIQEYFGAHGEBISEHKIKNFRKEMWNNFRKIMWAML----- 300
 QY 1449 IGAKRDDTEYGNVNVFSDKSTLLEPARKPOFLRMULTYDYDYCYTRQYKLVQOK 1508
 DB 301 ---SEKKNIN--NCKNIPQEL---QITOWIKENHGEFLERDRSRSLPKRSK 345
 QY 1509 CKSNDOLK-CDECKKCKCEDYKY-MKKKKEMIPQDKYKXKRRDRKRRDRHIGWVYDY 1566
 DB 346 CKNNLTYACECECLDPCMKTRDWTIKSKFEWHITSKETQKYKE----- 392
 QY 1567 TGTNATDYLNRFTASCGDCKRGSASVVOIRNTOLEKQAYYADKCGC---TKFIE--- 1619
 DB 393 ---NAENYL-IKISEKNKD-----AKVSLILNNDARESKYCDCKHTTTLVKSVIN 439

QY 1620 -ND---DKYTNISSKDKCK-GLVKEA-NTGALKQKQNGKPNNTNLIKELTEDEVLPSPRL 1672
 DB 440 GNDNTIKERREHIDLDDESKFQCKNSVDNTKRWCKNPLYLS-----TKDVCVPPRQ 494
 QY 1673 RICEFALDGNVTDPEYKDEGL--RRRLMEVATGEGYNGQYKKEKKEKIKTSDAHKY 1730
 DB 495 ELCIGNIDRIY-----DKNLMIKHHILAIAYESRILKRYKKNKDKDE----- 538
 QY 1731 STEVPPCSAMKSFYDLRIIIGIDNLEDEKQKTEENLKKIKNKGTSYKGSSTGTGP 1790
 DB 539 ---VCKLIINTFADIRIDIGITDYWD---LSNKKELKQINTNSKYVHRKKN----- 585
 QY 1791 GSTARKFEWNNKCEYWNAMIGYKRGDRGDSNGSARSDEJLKKGCSVPDDDYPMGN 1850
 DB 586 KLFREDEMKKVIKKDVMYNNISVFK-----DKTYCKE 617
 QY 1851 RDEGTAOPLRMPALWEGDECKHKEKELVAGANDYTGDNEDKRRKCTDCTQYKFF 1910
 DB 618 DDIENIPQFFRWFSEMGDDYCODKTKMIEPLLKVECKEKPCEDD---NCKSKSYKEM 672
 QY 1911 ISEMPQYKQIKKYE---NKDKIYSHPAKDAEDAREYLDQKIKCENKSGDCEY 1966
 DB 673 ISKKEEYKQAKQYQYQKGNMYKYSFKSIKREYVILKYSKX-----CSNLFEDEF 727
 QY 1967 K-----C--MKDVSTORLTDGNSQNMPSASLDDEPEY---EGKCNQVPRG 2007
 DB 728 KEELSDYKKNKTCMPKVDVPIIS-IIRNEDTSEANPEENTETLAHTEPISIEGKG 786
 QY 2008 PPRVRETPSPRVSILSKATASKKAKTAPPTQPKKYENLTENKRAOTRRRAAQOTRK 2067
 DB 787 NQOKRRDSDS---LSKISVS-----PENSRRP-----ETDAK 814
 QY 2068 RSTATTESDVG--IMVAILISNKPDSRGIEGCPKTYGQYPRKGCYKSKENENIC 2126
 DB 815 DTSNLKLKGDVDISMPRAVIGSSPN-----DNIN-----VTQGDNDISGV- 855
 QY 2127 MPRRRKCLINNIOYLYNETENKRDNDIKFAIFKCALETQFLMLKTYIENPAENELON 2186
 DB 856 ---NSKPLSDVRDRDKKELEDQNSDESEYV-----NHLSKSPSTINGDS 899
 QY 2187 GTIPDEFKRIPTYGYKDMFEGTDISNDKKIITVNSVTTLNENKKODKKRDEEL 2246
 DB 900 GS-----GSATVSESSSSNTGLSIDDRNGDFVFTQDANTEDVIRKENADKDEDE 951
 QY 2247 RFIWEKNNKFIWEGMIGLYHLDEN---EKEKIDNTQYNDMTLFTSLEFYRRP 2302
 DB 952 KGADER---HSTSELSPEEKMLTDNCGNSLN-----HEVAKH 990
 QY 2303 QFLRMFTWAEFCNKRREOLK-----LEAGCKEYBCNGSN----- 2339
 DB 991 TNSNDVQOOSGIYVMNYEKELDTLENPSSSLDEGKAHELSERNLSSDQMSNTPPL 1050
 QY 2340 DGKQOCEACVYQNFYKKKKTEYEROREKPKD-----KDGAKYKQDPSTER 2388
 DB 1051 DNTSEETTER---ISNNEYKVEREDERTLTREYEDIVLKSMMNRESDDGELYDE---NS 1104
 QY 2389 DIEKATCAHEYLNNKIKELCNKDCSQKQSPSOLPKTQOQSQSDANDMESID---Y 2444
 DB 1105 DLSTVNDSEDAEAKMK---GN-DTSEKSHNSOHIEBDOOK-----NDMTVGDLDGTH 1155
 QY 2445 VPEEFNKECEPELSK---KSMIHTKRTIPEKPIPNVCYKAAAYLSKEANNMDITLK 2499
 DB 1156 VONEISVYVTEIDEKELRESKSIKH--KAEERLSHTDIKH-----INPEDRNNTLHLK 1209
 QY 2500 EKFLPIESTKESKSNSTNNN 2521
 DB 1210 D-----IRNEENRHLTJON 1224

RESULT 13
 US-08-487-826B-4
 ; Sequence 4, Application US/08487826B

```

Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Slim, Kim L.
APPLICANT: Chitlur, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-Zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-4
Query Match 3.2%; Score 617; DB 2; Length 1435;
Best Local Similarity 21.8%; Pred. No. 3.9e-35;
Matches 332; Conservative 215; Mismatches 557; Indels 418; Gaps 69;
DB 1101 EACNEVDIIGLMKEQ--NIISDKYKELHEQAOQSVNSGIEASSTARNHIDRVYIEF 1157
20 KAREY-DI--KENEFIDYKKEFENLDKKKYGNYOKT-----DKRIFTF 62
1158 L-SRLVQONGGSKMS-----GTSDSAVIGTNTTYENVAGVLAHDYGNFDDCQSQNECDE 1212
63 IENKLDILNNSKFNKRSKSTGTPDN--IDKNSLIN----- 96
1213 KSDGKDEKAFYRDKPODHGACGCGSGSPTRVQ-ITKKAERKDETECKVNDILKEN 1271
97 ----KHNEMEFNNNYSFLSTSLIKONKYVPINAVRSRILSLFDSR--INNGANTS 149
1272 DGKQVEDCHPKKNSNGYPRMOCGINLVEDPRVCMPPROKLCVHLANDNEIKLQSQ 1331
150 SNNEVLSNCEKRRG--MMWDCRKN-DRSNVYCIPIRRIQICIVLS--TIKTYTK 201
1332 VNLEAFIKSAAEFTFSPWYYSKDEGNELDKELKRIIPAFILSMYTFEDYDPL 1391
202 ETKMDHTEISKKE--SOLLKKNKKYNS-----KFCNDLKNSEFLDYGHIA 246
1392 FTGDISGHHGSGKLEQIDSLFENGDKSPNGKT--ROEMWTEHSEIWEAMICALVK 1448
247 MGNMDPG-GYSTKAENKIDGVFGANGELISEHKIKNFRKEMWNEFEKILMEANL----- 300

```

```

QY 1449 IGAKKDDPTENYGYNNYKFSDKSTLLEFAKRPQFLMLTWMYDDCYCROKYLKVOEK 1508
DB 301 -----SEHKNNIN-NCKNIPQEL-----QITQWIKEMGEFLEEDNSKLPKSK 345
QY 1509 CKSNDOILK-CDTECNKKCEYVYK-MKKKEWLPQDKYVYDERDKKRPQHGVMVTDY 1566
DB 346 KNNVTLFEACEKECIDCMYRMWIIKRSKFEWHLISKEYETQYVYKPK----- 392
QY 1567 TGNATDYLNRKFTFASCGDPRGSASVYQRIOLLEKAYVADAKHCC--TKFIE---- 1619
DB 393 ----NAENVL-IKISEKND-----AAVSLINNCDAEYKICYDCKKTTLLVKSVLN 439
QY 1620 -ND--DKYTNISSDKCK-GLVKEA-WTGAIKWNKGPNYNNILKELTEDVLYEPSRL 1672
DB 440 GNDTIKEKREHIDLDGFSKFGCDKNSVDINTKWECKNPIYLS-----TKDVCVPRRQ 494
QY 1673 RICEHALDNGVTPYVDEGL--RKLMBVAATEGINLGOYKKEKEREKITSDAHKY 1730
DB 495 ELICGNIDRIY-----DKNLMILKEHILAIYESHILKRYKKNKDKKE----- 538
QY 1731 SYEVPSCAMKSPFYDLRIILGIDNEDEKOKTEEMIKTFNKGTSYVSGSDSTGNP 1790
DB 539 ----VKIINKTPADIRDIIGTDYND--LSNRKLVGKINTNSKYVIRNNKN----- 585
QY 1791 GSTARKEFFNENKECYVNNAMICGYKRGDRDONSARSODELKKCGSVSDDYPMGKN 1850
DB 586 DKLEFRDEMMKYIKKDVNVNLSWYK-----DKTYCKE 617
QY 1851 RDEGTAVQFLRWAPGEGEDCKHKELEKLVGACNDYTGDNEDKRRKCTDACTQYKPF 1910
DB 618 DDIENIPQFPRWSEWDDYCODTKMIELTKVCKKPEED--NKSCKNSKSG 672
QY 1911 ISEWKPOYEKOIKRYGE--NKDIYSEHPVAKDAREYLDKQIKCIGENSGCEY 1966
DB 673 ISKKEEYNQAKOYQOYQGNNTKMTSEKSIKPEYLLKYSK-----CSNLFEDER 727
QY 1967 K-----C-MKDVSTQRLDNGSNMPASLDEPKREV--BGKNCQVPRG 2007
DB 728 KEELSDYKKNKCTCPFKVDYPIIS-IRNNEQTSQEAIVPEENETIAHRTPEPISSEPKG 766
QY 2008 PPRVRRPTPRVSLISKATASKREAKTAPPTQPKVENLITMKRQOTRRRAOQTRK 2067
DB 787 NEQKERDDDS--LSKISVS-----PENSRP-----ETDAK 814
QY 2068 RTSTATTESDVG-TWYKAILSNKPDGSGIEGCPRTYQYPRKGCYVSKENENGIC 2126
DB 815 DTSMLLKLKGDVIDSMKAVIGSSPN-----DNIN-----VYEGDNISGV- 855
QY 2127 MPRRRKLCINNIQYLYETENKRDNDIKAEFIKCAAIETQFLMKYTIENPAENELON 2186
DB 856 ----NSRPLSDVDYPRDKKELEDQNSDESEYV-----NHISKPSINNGDS 899
QY 2187 GTTDFEKRIMYIYGYXKMFEGTIDISNDKITTIVNSVTTILNENKKKOKKOKKDEEL 2246
DB 900 GS-----GSATVSSSSSNTGLSIDDDNGDTEVYRTDTANTDEVIRKENNDKODE 951
QY 2247 RKIFMERKKFIWEGMILGYLHLDEN--EKEKIRDYQYQYDMTKLTPSLEEFYKRP 2302
DB 952 KGADDER-----HSTSESLSPEKMLTLDNGGSLN--HEBYKEH 990
QY 2303 QFLRWFTMAEFCNKRKEOLK-----LEAGKEYECNGSN----- 2339
DB 991 TSNSDNYOQSGGIYNNMVEKELKJLTLENPSSSLDGFAHEBLSPLNSPDQMSNTGPL 1050
QY 2340 DGKTOEACACVYQNFIFKMKTKTEYEROREKFD-----XDGKKYKQYPTER 2388
DB 1051 DNTSEETTER--ISNNEYKNERDERLTKYEDYILKSHMNRSDGELYDE--NS 1104
QY 2389 DIEKATCAHEYLNNKLKELGKNKDCSCMORPSSQLPTTQOSQSSDANDMESPESID--Y 2444
DB 1105 DLTSTVNDESEDAEAKMK--GN-DTSEMSHNSQHISSDQK-----NDMKTYGDDLTTH 1155

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:32:15 ; Search time 79.1271 Seconds

(without alignments)
3586.886 Million cell updates/sec

Title: US-10-087-013-2

Perfect score: 19407

Sequence: 1 MFSCKRYTIKGNASSLE.....IDLINDSLVILLYMKY 3542

Scoring table: BLOSUM62

Searched: 301932 seqs, 80129803 residues

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications, AA.*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	2972.5	15.3	2710	9	US-10-153-273-12
2	1032.5	5.3	700	9	US-10-153-273-10
3	665.5	3.4	1086	10	US-09-924-154-15
4	633	3.3	921	9	US-10-153-273-8
5	617	3.2	1435	9	US-10-153-273-4
6	599.5	3.1	1421	10	US-09-924-154-13
7	594.5	3.1	1501	10	US-09-924-154-17
8	591.5	3.0	1143	10	US-09-924-154-14
9	431	2.2	749	9	US-10-153-273-6
10	382.5	2.0	972	10	US-09-924-154-16
11	365	1.9	411	9	US-10-153-273-20
12	343.5	1.8	1115	9	US-10-153-273-2
13	291	1.5	6281	10	US-09-815-242-12996
14	289	1.5	1639	9	US-10-087-464-10
15	287.5	1.5	2125	10	US-09-919-172-29
16	279	1.4	362	9	US-10-153-273-18
17	279	1.4	2789	10	US-09-801-574-57
18	276.5	1.4	2285	10	US-09-933-183A-2
19	276	1.4	411	9	US-10-153-273-19

20	273.5	1.4	311	9	US-10-153-273-21	Sequence 21, Appl
21	256	1.3	1605	9	US-09-820-843A-73	Sequence 73, Appl
22	234	1.2	2478	10	US-09-815-242-5816	Sequence 5816, Ap
23	234	1.2	2478	10	US-09-815-242-12967	Sequence 12967, A
24	230.5	1.2	3158	10	US-09-815-242-12611	Sequence 12611, A
25	225.5	1.2	2139	9	US-10-023-219-4	Sequence 4, Appl1
26	225.5	1.2	2139	10	US-09-727-384-6	Sequence 6, Appl1
27	223	1.1	2785	10	US-09-801-573-8	Sequence 17, Appl
28	218.5	1.1	324	9	US-10-153-273-17	Sequence 50, Appl
29	215	1.1	1331	9	US-10-087-464-50	Sequence 5703, Ap
30	215	1.1	2025	10	US-09-815-242-5703	Sequence 108, App
31	214	1.1	807	9	US-09-820-843A-108	Sequence 5635, Ap
32	213	1.1	2368	10	US-09-815-242-5535	Sequence 12389, A
33	213	1.1	2368	10	US-09-815-242-12389	Sequence 7, Appl1
34	206	1.1	1400	10	US-09-764-176-7	Sequence 1, Appl1
35	206	1.1	1404	10	US-09-811-045A-1	Sequence 493, App
36	201.5	1.0	1095	9	US-10-076-622-493	Sequence 493, App
37	201.5	1.0	1095	12	US-10-007-805-493	Sequence 565, App
38	200	1.0	1341	9	US-10-076-622-565	Sequence 573, App
39	200	1.0	1341	12	US-10-007-805-565	Sequence 5834, Ap
40	200	1.0	1349	9	US-10-076-622-573	Sequence 553, App
41	200	1.0	1349	12	US-10-007-805-573	Sequence 553, App
42	199	1.0	2437	10	US-09-815-242-5834	Sequence 475, App
43	198.5	1.0	1013	9	US-10-076-622-553	
44	198.5	1.0	1013	12	US-10-007-805-553	
45	197.5	1.0	1002	9	US-10-076-622-475	

ALIGNMENTS

RESULT 1
US-10-153-273-12
Sequence 12, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 12
US-10-153-273-12

Query Match 15.3%, Score 2972.5, DB 9, Length 2710;
Best Local Similarity 26.5%, Pred. No. 2,6e-155;
Matches 865; Conservative 442; Mismatches 873; Indels 1083; Gaps 120;

QY 9 IIRGNASSLEEDAKSPIIKESHKSARAVLEARYAKIRHPSK-YAKEHVDLKGDTJTK 67
Db 6 LAKMGPREAGGDIED-----ESAKHMFIDIGDVYKAYEKEKRGKGLQRLSEA 58
QY 68 EFRGSPSTPVNKNINYYPPYPCNDIHEHTNLRDYDNLHHPCHGRQNRFEDESEEC-G 126
59 KFEKNESDPQTPBD-----PCDDHRYHTNV---TTNVINPCADRSVDRESDYGGQCTH 110
QY 127 NKTIRYKRNDAIACAPRRRHMCNKLALNDINTQNHIDLGVLVYAKYGESEIVNN 186
Db 111 NRIKDSQGDGNKACAPRYRLHVCODNLEIEPIKITNTNHLVYDCAKKEGOSITOD 170
QY 187 HP-HKGT-----SDACTALARSFADIGDIYRGIDMFKPVNH-----KYETGLREYFKKI 235
Db 171 YPKYATYSGDSPSQICTMLARSFADIGDIYRGIDYLGMPQELKQROQLNNKITFGKI 230
QY 236 HD---GMEDEVKNDYPDGSGNYKLEBAMNVRNKKVYEAITCDASTYSGTFMOSESNT 292
Db 231 YKLNLAARARYGND--PE---FFKLREDMWAIANRETVAKITCA-NWNTYF-HATCMR 282
QY 293 PLESNKCGHKGKVPNTLDYVPOYLRFDEWGESECRKRNITKLVKDCR---NDKE 348
Db 283 GERTKCYCRNDQVPTDYDYPOYLRFWEEMADECRKNNKIDYKRNCGKDEKED 342
QY 349 RLYCSINGHDCITTYIKKGIHLDMKCTDCSTKCAVEYWLGNQDPAFKREKYEKETO 408
Db 343 R-YCSHNGYDCETKRAIKRLYGRKQICSLACNPYVWMINNKEOEFKQKKYDEIK 401
QY 409 SY-----LSNDKRFVNNINSE-YKQFYEKLEKQYATNDPFLNLTNGKXC--- 454
Db 402 KYENGASGSGRQKRAAGCTTTNYGCEKFFDELNKSSTYKIVDFLEKLSHEEICTYAK 461
QY 455 --KGLPEKIDITFTNSA-----DNGGIFRYSEYCOVPCDGYKCDGIKYTH-KSPND 504
463 DEEGGIDTFKKNVNSDTSASGTSVSGSTFYRSKYCPGCPYGVK---KVNNGSSNE 517
QY 505 RERVNN-----EDKPPMGVAPNTIYLYSGNEGODITQKLENCN-----SS 547
Db 518 WEKRNNGCKSKLYEPKDEGTTITLIKSGHDIIEKLNKFCDERNGDTINSGG 577
QY 548 TNYKDKNN-----QKWECCYKDEINRCKLEONTIINNBPK-----II----- 586
Db 578 TGGSGGSGNGRQELYEEMKC-YKGEDVYKVGDEDEDEDEYENKAGGLCLIKNOKKAKE 636
QY 587 -----SEINFEWLYTYLLRDTIKWDLKTCI-NNTTHC-IDECNHC 629
Db 637 EGGNTSEKEPELQKTPFPFYYVAHMLKDSIHKKKILQGLCLQNGNRKGNKKNCCND 696
QY 630 LCFDWWOKKEEEMNSIKLFTKKKNIO-----QSYSSINNLFGYF--- 672
Db 697 ECFKRWITQKDEMGKIVQHP-KTONIKRGSGSDNTAELLIPFDHLYQVLMJOEEFLKGD 755
QY 673 -----FYMDKLDKDEAKKELMENIRKKRNFSENLENNPD-----YLENAITELL 718
Db 756 SEDASEEKSSENSDAEER-----EELKHLREIIESENNOBASVGGVYTOCKINIMDL 809
QY 719 DHKETAIC-----KDNNTNEACETSHNATTPCVKPRGGTO-PTKNIKEIAQ 766

Db 810 NYEKDEADLCLEIHEDEEEKEKGDGNECTIEGENFRYPCCSGSGNRRYPVLANKVAYQ 869
QY 767 YFKRSAYEEDARRNGIHLKKAHEGITYKRGRRKDFKDMICRIMIKHSNRNLGFSNGPCD 826
Db 870 MHHKATQLASRAGRSLGDISLAQFKNGRNGSTLKGICIKINENYSNDSNGSGCPT 929
QY 827 GKGTDGIOTRFVYGTGEMVEVDEPHMKKHEDYIMPPRRHITSNLEHLQTDHPLNGNI 886
Db 930 GK-DGDHGVRRNRHIGTEMNIEGKQTSYKNVFLPRRRHMCTSNLENDV-----GSV 982
QY 887 V-DDLVNNSFLDVLISAKYEANKIIRMYKEKNKNGPREVDP---KHQITCAIRYS 942
Db 983 TKNDKASHSLDGYOAAKTAADAELIKRKQDNNI-----QLTDPQKDQEMCAVAYS 1038
QY 943 FADIGDIIRGRBLERNNGMVAQLQGLETVGNHKS LG-KGNDKYND--APRYLKL 998
Db 1039 FADIGDIIRGRMWDKEDSSTDMETRLITVFNKIEKHGDIKDNKRYGDSKRPAYKKL 1098
QY 999 RENWNEANRAKYWEAMKCDIKYLDKSGHOSSTSSGYSDHTPLDDYIPQKLRMTEMA 1038
Db 1099 RADWNEANRHQVWAMKC-----ATKGLIC---PQMPVDDYIPQRLRMTEMA 1143
QY 1059 EWCYVQKKEVDKLEKCECKDNQDGGCTKESGTGCTKEACNEYNDIILGKREQWN 1118
Db 1144 EWCYCAQSQEYDKLTKICADCKSKGDG-CT-QGDVDCGKCAACDKYKEIEKMEQMR 1201
QY 1119 IISDKYKELHQAOQSVNSGIEASTAKNHID--RNVIEFLSELVQONGK---SNK 1171
Db 1202 KISDYNLILYQAKTTSNPG---RTVLGDDDPDYQDMVFLPIRHASIAARVLYKRA 1257
QY 1172 SGTDSASAVIGTNTYEVNAGYVILHDTGNFDDQSQSNEFCDEK-----SDCKDNKXAFR 1225
Db 1258 AGSPTEIAAARPIPYSTAGYIHQELGYGCGQDQCEKHKHATSTSTKEKKEITFK 1317
QY 1226 DKPOHDACCGCKSGSKSTRQVQITKKAABEEDTECTVNDILKENDKQVDECHPRKN 1285
Db 1318 QPPPEYATACQKINSQTE---EPKKKEENVEASACKIVEIKELGKNRFTVGCNPKES 1373
QY 1286 SNGYPMQC-GNINLVEBPVRCMPPRROKLCVHLANDNEIKLQSOVNLKEATIKSAAA 1344
Db 1374 ---YPMDCRKNNDISHD-GACMPPRROKLCVLAHSEQENIKTNDNLKDAIKTAAA 1429
QY 1345 ETFPSWYIYKSK-DGEGNELDKELKEGKIPAPLRSMEYTFGDRPLFTGDISKGHEG 1403
Db 1430 ETFPSWQYIYKSKNSSEAKILDR---GLIPSQFLRSMTYFGDRDCLMTDISKKNQNDV 1485
QY 1404 SKLEQIDSLFKNDQSPNGKTRQEWHTSHSHIEMALCALYKIGAKKD--FTENTY 1460
Db 1486 AKADKIGKFEFSKDGSKSPSGLSQEWKTINGPEIMKGMICALTKYVTDNKRKIKINDY 1545
QY 1461 GYNNVKEFSK-STLLIEFAPKRPQFLRMLEWYDYCYTRQYLDKDVQEKCS-NDQKCD 1518
Db 1546 SYDRVNOSONONPSLEFPAKPOFLRMIMGEFCAERKQKKEIKITIDACNEINSQOQN 1605
QY 1519 T---ECKKCEDIYKAK-KKKEWIPQDKYTKDERDKRPRDROHIGVMTDYGTATADY 1574
Db 1606 DAKHRCQACRAQOYEYENKKEFSQGTNNFVLKANVQPDPEKGYEKD----- 1656
QY 1575 LNRFTASGDKPGPSASVQVQNLQLEKQAYYDADHCGCTKLTENDDKTINISKDKCK 1634
Db 1657 -----GVQF-----IQGNEYLLQK---CDNNK-CSC----- 1678
QY 1635 GLVEANTGAIKMONKGPNNYNNILKELTEDVLPSPRLRICFHALDQNTYDPEKENG 1694
Db 1679 -----MGNNVLSPKPK--- 1691
QY 1695 KRLMEVAAEGYNLGOYTKKEKEKIKTSDAHKSYEVPSCAMKYSFYDLRDIILGI 1754
Db 1692 -----PFGKY-----AHKYP----- 1701
QY 1755 DNLDEKQKTEENLKIFNKNGTSVKGSQSTJGNGSTARKPFWMENKECYWNNAMIGY 1814
Db 1702 ----- 1701

QY 1815 KRGDDGNSGNSARDEDLKCGSVSDDDYPMGNRDEGTAVQFLRMFAWGEDECKRH 1874
 Db 1702 ----- 1701
 QY 1875 EKELELVGACNDYTCGDNEDEKRRKCTDCTOYKFFISEMKPOYEKOIKYGENKDIYS 1934
 Db 1702 ----- 1704
 QY 1935 EHPVAKDAEDAREYLDKOLKICENKSGDCEYKCMKDVSTQRLDGNQSNMPASLDEPK 1994
 Db 1705 ----- 1706
 QY 1995 EYEGKNCVPRGPRVRRBTPSPVSLISKATASKREAKTAPPTOPKRVENLTTEMRA 2054
 Db 1707 -YOGK---HVPSPIP----- 1725
 QY 2055 QFTFRRAAQTTRKRTSTATTESDVTGMVAKILSNKPDNSGIEGNPKTYGQY--PKWG 2112
 1726 -----PEAPTYVIVGCSIVKTLK---DTNFSDACGLK-YGKTAPSSWK 1766
 2113 CI-----VGKSENGEICMPRRKRLCINNIOYLANET-----ENKRD 2151
 1767 CIPSDTKSGAGATGKSGSDSGSICIPRRRLIYVGKLO--EWATALPOEGAAPSISRA 1824
 QY 2152 NDIKEAFICAIETQFLMLKIIE--NPAENE-----LONGT 2188
 Db 1825 DDLRNAFIOASAIETFLMDRYKEKKPOGDSQOALSOLTSTYSDEDEPPDKLONK 1884
 QY 2189 IPDEFIRIYTYGDKKDEFFGDISNDKIIYVNSVTTILNENKKKODKDELRK 2248
 Db 1885 IPDFLRLMEYTYLDYRDILVHGNTSDSGTNGSNMNLVLEASGNKEMOKIOEIQ 1944
 QY 2249 I-----FEKNKRFIEMGIYGLTY-----HLTDNEKEKIR 2280
 Db 1945 ILPKNGFPLVPKSSAOTPKMWNHAEISWKMICALTYTEKNPDTSAAGDENKLEKD 2004
 QY 2281 DNT-----QYN-DMTRL-----TPSLEEFYKRP 2302
 Db 2005 EYVEKEFGSTADKRGVAPPTGYTKOYDEKYLEDTSGAKTPSPASDPPLSDVFLRP 2064
 QY 2303 QFLRMTEMAEEFCNRKBDLKLKLAGCKEY-----CN---GSN 2339
 Db 2065 PYRITLSEMGONCKKRRKHLKQIKHECKVEENGSGSRGCGITROYSGDEACNEMLPKN 2124
 QY 2340 DGKTOE-----CAECVYQNFITKWKTEYEROREKFKDKD----- 2376
 2125 DGVTPDLKESCAKPCSSYKRWIESKGEKEQEKAYEQDKDCVNGSNKNDHGFCELT 2184
 2377 -GKKYDY-----PSTERDEKATCAHEYILNMLKELCGKDCSCMOKPSSQLPKTTQOS 2430
 Db 2185 TTSKARDFLKTLGPKCPNNVEGKTIFFDDDKTFK---HTKDDPCCLKFVGNCK--- 2234
 QY 2431 QSSDANDMPESLDYVEEFKCEPSELKSGSMIHTKITEPKIPMNCVKAAYLSKEA 2490
 Db 2235 -----DECD---NSKGTCCRNK-----NSID-----ATDI 2256
 QY 2491 ENNMIDTLKKEFLPIESTKEKESKNSMTNNPCDPKPYADPYIGRRNCPEN----- 2543
 Db 2257 ENGVDSTY-----LEMRVASDSKSGFNGD-----GLENNCRAGIFEI 2295
 QY 2544 REENRRKVDYEMKCYKNSKTYQEKRRVCYPPRRHMLCLNLDLKIERLKSNTYLLKMR 2603
 Db 2296 RKD-----EMKC-RNVCYV---VCKP----- 2313
 QY 2604 RTARNEGIDILKMFNSENGCANMPIDTMKYSFADLGDIVRGTDMLRIGGYLPVVEIKLY 2663
 Db 2314 -----ENVNGE-----AKGHHIOTIALVR---RWY 2336
 QY 2664 KVEEYLYGKRNKNGKRNKYND-----VQTFRSAMDMANRDKIMAKMTCKAPEDAKFERKG 2719
 Db 2337 EYFEEDYNKIKHKISHIKNGEISPCIKNCVEKVMQDKRRE-WKELTERKD----- 2387

QY 2720 RMDGEFIRILLQCKGKHDDPVVDYIIPORFRMTWSEYCCALMELEKFRKSCDHCK 2779
 Db 2388 -----QYKNDNSDDDNVRSFLETLIPQITDAN---AKNKYIKLSKFCNSCG-CS 2432
 QY 2780 TSDRCKNDVDENKCEQCKTKCOEYKPNVLLKMKSLFLOSKYKVELBOPIYTIISYIDH 2839
 Db 2433 AS---ANEONKN-----GEYKDAI-----DCMLKRLD-----KIGCECK 2465
 QY 2840 QNFVOKLTKFESCSVSEFSEYELTSCIKLYKFENEDSSNIRTYAFETPKSYKE--- 2896
 Db 2466 HH-----QTSDETCSDTPQQTLED--ETLDDDIETBEAKKNMMPKICEVYLKTAQODE 2518
 QY 2897 -ACSTLPKRNPLDNCPTQONKD 2918
 Db 2519 GGC---VPAENSEEPAATDSGKE 2538

RESULT 2

US-10-153-273-10
 ; Sequence 10, Application US/10153273
 ; Patent No. US20020169305A1
 ; GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
 Chitnis, Chetan
 Miller, Louis H.
 Peterson, David S.
 Su, Xin-zhaun
 Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660

COMPUTER READABLE FORM:
 MEDIUM TYPE: floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/153,273
 FILING DATE: 21-May-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:
 LENGTH: 700 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 10:
 US-10-153-273-10

Query Match 5.3%; Score 1032.5; DB 9; Length 700;
 Best Local Similarity 34.8%; Pred. No. 4.6e-49;
 Matches 247; Conservative 110; Mismatches 250; Indels 103; Gaps 22;

QY 136 NDATACPRRRRHCNDKLEALNDINTONIHLLGNLVTAKEGESIVNNH--PHKGT- 192
 Db 5 NKVACAPRYRLHLCYDNLESID--TSTTHKLLLEVCAMAKEGNSI-NHTHTOHRTN 61
 QY 193 -----SDACTLARBFAIDIGIVGIDMF-----KPNVHDKVETGLREVFVKIHDGMEDE 242
 Db 62 EDSASQCLCTVLAKEFADIGIVRGKIDLYLGYDNKNEKORRKLQKIDIKFKIHT--KDV 118
 QY 243 VKND-----YNPDG-SGNYYKLRBAMNNVNRNRYWEAITGDASYKSGYFMOSSNTPLES 236
 Db 119 MKTNGAQERITIDAKGDFEQLREDMTSNREYVWKALICHAKREAFYFKIACNNGKGT 178
 QY 297 NPKGHAQGVPTNLDVPOYLEKWFDEWGEFRCRKNIKLKYKDSQRNDKERLYCSHNS 356
 Db 179 NGCHCIGSGVPTFYDVPQYLRWFEEWADFCRKKKLENILOKOCROVEQMLYCSGNS 238
 QY 357 HDCTTTIKKGIHLNDKCDGCTCKCFVFWYMGNOEAFKOKREKYEKEI----- 407
 Db 239 YDCTKITIKKGIHLNDKCDGCTCKCFVFWYMGNOEAFKOKREKYEKEI----- 407
 QY 408 -----OSYLSNDKRFVNNINSEYKQFYEKLETOYATNDTFLNLNEGKCYCKG-- 456
 Db 299 KRTKRAARSSSSDDN-----GYESKFFYKLLKEVGYODVDKFLKILNKEGICQKOPQ 350
 QY 457 -GLPGKIDITFTMSADDKGIFYSEYCYOCPGVCDCGKITKTHKSDNDRERYNNEDYRP 515
 Db 351 VNGEKADNVDFTEKTYK--TFSTHTEICEPCWGLKGGPRPVKDKTCSAKTWTYDP 409
 QY 516 PMGVKPTNTVLYSGNEQDITOKLENFCSNTNYKDNKNKQWECYKRD-----ENINCG 570
 Db 410 K---NITDIPVLYPRKSOQNLKRYKNFCEKGP--GGGQIKKQCYDEHRSSKNNNC 465
 QY 571 KLEQNEINNDNPKIISFNFEFLWYTYLLRDTIKWMDLKTCINNTT---HCIDECN 626
 Db 466 VESTWMDTQOGQOTVKSIVYFWMVHMDLHDSVEKTELSKICINNNTNGTCRNNNKCK 525
 QY 627 RNCICFDRWVOKKEEEMNSIKKLETKKKNIOOYSYNNINLEGEFFYKMDLDEAKM 686
 Db 526 TJOCGCGQKVEKQOEMAKIDHFGQDIIYQO-----KGLVFSRYGLDLV---LKG 576
 QY 687 KEMENIKRKKKKEFSNLENNRDYLE-----NAIELLDHLKETATICKDN 732
 Db 577 GNILONIKDVHGDITDIKIKILDEDAVAVVGLGKMDTIDKILQHEKBEAQCKOK- 635
 QY 733 TWEACE-----TSHNATNCPVPRGTOPTKNIKELAOTFKRSAYEE 775
 Db 636 -OECEKKAQOESRGRSAETREDERTQ--OPADSAGEVEEEDDDDYDE 681

DLT 3

US-09-924-154-15
 ; Sequence 15, Application US/09924154
 ; Patent No. US20020127241A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Narum, David L.
 ; APPLICANT: Sim, Kim L.
 ; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 ; FILE REFERENCE: 05213-0465 43170-262105
 ; CURRENT APPLICATION NUMBER: US/09/924,154
 ; PRIOR FILING DATE: 2001-08-07
 ; PRIOR APPLICATION NUMBER: US 60/223,525
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 15
 ; LENGTH: 1086
 ; TYPE: PRF
 ; ORGANISM: Mammalian
 US-09-924-154-15

Query Match 3.48; Score 665.5; DB 10; Length 1086;
 Best Local Similarity 21.08; Pred. No. 1,5e-28;

Matches 312; Conservative 186; Mismatches 431; Indels 555; Gaps 60;
 QY 2070 STRATTESVGTWVKAIIISNKDPSRGIEGNCNPKTYGQYPMGCIYVGSKENENGICMP 2129
 Db 7 SKSIIRNKDVADISIKPFENKFIIDYSLIECGKFI--KEMKICIDNCFK--SNLCAPI 62
 QY 2130 RRKIKLNNIYOIANTETEN-----KRD--NDIKFAFKCAIETQFLMKYITENPAEN 2182
 Db 63 RRIQCLIVNI--ILFSENEVEYKNDISNNKREKILAVKLESMLVQK-----N 113
 QY 2183 ELONGTIDPEFRIMYTYGYDKDMFEGTDSINDKILITVNTSVTTINE--NNKKQDK 2240
 Db 114 NEYNSKLCDDIR-----WSLGDYDIIIGRDLIKYNNTDYIKQFKIRINNEYNNELNDE 169
 QY 2241 -----KDEBLRILFWEKKNKFLWEGMIGLYLHLDENEKIRNDYONDTKLTP 2293
 Db 170 LNNELNDEKNIKLRKEWKEKEDIMEE-----TKHND--KRIE 208
 QY 2294 SLEEFV--RPOFLRWFTEWAEFCNKRKEOLIKLEAGKEYEBCNSNGKQOECACACT 2352
 Db 209 KCKYFAKDEPOLVRIKIEKSKOFLDEKNYMLFTLRNTYENMIHNN-----CKQ 259
 QY 2353 YONETKMKTEYERQREKFKDKDQK-----YKDYPTERDIEKATCAHEYLMKL 2404
 Db 260 YKKWYQNRKKEWTELSNEFNKIFPERNYQIHSNIFKEKYNNDYILGTLNTEYNN-- 317
 QY 2405 KELCGNKQSCQKORSQOLPPTQOSQSSDANDMPESLDYVPEEPNCEPELSKQSMI 2464
 Db 318 -----CKEKP-----ELVSAK----- 329
 QY 2465 HTKITEPRIPMNCVEKAAYILSKAENNDITLKEKPIESTREKESKSWTNPPCD 2524
 Db 330 YNLKAPNAKSP-----RIYKSKHEESSVFGKTKI----- 360
 QY 2525 PKKPYAPKTYGRNPPCNRBENRKYDYETKCYKNSKFTQOKKRVCPVPRHMCILNL 2584
 Db 361 -----SKYKKKWNCSNNK--VTKPBGVGPBRROOLCGIYI 395
 QY 2585 DEIK---IERLSDSNYLLKWRPTARNBGIDILINFSENGCANPICDMKYSFALGD 2641
 Db 396 FLINDGNEGLKDH-----INKAANYEAMHLKAKTENAG--DKTCMILISYALIGD 446
 QY 2642 IVRGTDMLRIGIYLPVEI---KLYKVEEYLYGKWRKNRKNRYNDYQTFRSAMANDR 2698
 Db 447 IVRGIDVWR-----DINTNKLSEKFQKLIPOGGSRRKQNDNNE--RKNWMEKOR 494
 QY 2699 KDIMKAM-----TKAPBDATLFRKGRMDGFERITLLIQDKCGHKDDPPVDYIPORF 2750
 Db 495 NLINSSWVKHILPKGKTCK-----RHNNFK-----IPQFL 524
 QY 2751 RMWTWSEYYCKALMELEKFKKSCDHCKTSDCKNDYDENKCEOCRCQOEYKNFLKW 2810
 Db 525 RMLKEMGDECEMGTEVQKLEKICENKNSK-----KCKMCSSTIEKTIKER 573
 QY 2811 KSLFDIOSNKYKELIEQPIYTKISTYDVONVQOKLTFKSECSVBSFSEYIHTSK-CL 2869
 Db 574 KNEYNLQSKRFDG--DKLKNKNNLYN-----KFDKSAIYLSRESKQCS 615
 QY 2870 NYKFNENDGSSNIRTYAFEEYTPKSYKACS--CTLPSKN-----PLDN--CPTQDNKQGCHE 2922
 Db 616 NIEBND-----ETFTF--PKRYKACAVCENPSSKALKTITNVPIEESKR--SE 663
 QY 2923 LQFTFCSKNDYNNLDNNMNAVLYVNSSDDNKGVLIPRRRHLCPTPIAYNVRKDKKEI 2982
 Db 664 LSSLITDKSKN-----TPNSSGGG-----NY--GDROI 688
 QY 2983 LKK-----KLITSAFSQGLLGQIKYSEELC-----FEAMKSYADYDIIKGTDM 3029
 Db 689 SKRDVYHHGPKKEVKGGEKPRIDAIVATENTFTSNRNDIBKESKSDHSPVASKDI 748
 QY 3030 MDTSLSEKIKLIFETSNEATEENKRTWMENNRRQIWHAMLCGYIANSKYTLDBGMOQLPK 3089
 Db 749 -----KNEEPQRYVSEN-----LPK 763

```

QY 3090 DEETNOPLRLIMAKACKKRVSDSLTKCPKRSNEDNEFEASSELLROFGCONDIRKTY 3149
DB 764 IE-----EKMSSDSIP-----THIEA-----EENYOSYK 3191
QY 3150 SLNLILKNTMENTINIKYKOLKDOSSGJNDKPS-----EENYOSYK 3191
DB 783 -----KGOSSNSDNDPAVVGRESKDVNLHTSERIKENEKGVIK 822
QY 3192 SKDSOCALF-----NDINEIVTGTKNNE-----FKEYLKLYPGLYVED 3235
DB 823 TDSSKSIIEISKIPSDONNHSDLSQANEDSNCKETINPSTENKLIKIH-----874
QY 3236 ETKKNHVLGNINEEQYAPKALFFTPVDSFYQAPLFSTRVAYQYDPKNDILASS-- 3293
DB 875 --KTSDSDDHSGKIKSEIEPKELTESPLTK-----KTESAIGDKNHESVXSADI 925
QY 3294 -----ISVIVY-----SALGLIALHPMKKKRKSVDLRLININQOGEVMP 3334
DB 926 FQSEIHSNDNRIVSESVVODSSGSMSTESTRTDNKDKTSEDIAPSIN-----976
QY 3335 TLESKNRYIPRSGPKTYIEMGDTSGDEDKYMWDLSSDITSESEYEELDINDY 3394
DB 977 -----GHEKIGSSADDRGSEDSKSIIDKSENFENKSSHSDIKQSD-- 1017
QY 3395 VPGSPYKTLIEVYLEPSKRDIPS-----DDTPSNDTPKTN 3430
DB 1018 NEGSTDYESLTE--ESPKGDLSEVSPPSIDMLKPKKSPVTS 1058

```

RESULT 4
US-10-153-273-8
Sequence 8, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.

Chitnals, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273

FILING DATE: 21-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael

REGISTRATION NUMBER: 36,516

REFERENCE/DOCKET NUMBER: NIH21.1FMDV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8530

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids

```

? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? MOLECULE TYPE: protein
? HYPOTHETICAL: NO
? ORIGINAL SOURCE:
? ORGANISM: Plasmodium falciparum
? SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-273-8

```

Query Match 3.3%; Score 633; DB 9; Length 921;
Best local Similarity 27.8%; Pred. No. 7.3e-27;
Matches 224; Conservative 113; Mismatches 264; Indels 206; Gaps 42;

```

QY 948 DIIRGDLWRNDGMYLQGLHETVFGNIHRSLSKGNDRYN-----DDAPKYLKLENNY 1002
DB 80 DYITASDI-ENGNSI-----GNIDMYVSDKRNANFNGLDAGSANIRKGIKEQ 128
QY 1003 WEANRAKVAEMAKCDIKYLDKSGHOSYSSYCGYSDHPPLDDYIPQKLRMTWMAEWYC 1062
DB 129 MKC--AKV-----GGLDVCGLKNGN-----GSIDKQKQIILIRAL--LKRWEYFL 171
QY 1063 KYQKREYDKLEKCKECKDKONGGCKESGTCGTCTACNEYNIDILGKMGQNTISD 1122
DB 172 -----EDINKINAKISHCTKADNNESTCTNDCKPCKT-CVE-----EWINKRTWKNIKK 220
QY 1123 KYRELHE-----QADMYSN-----SGIEA-----SSTAKNH 1149
DB 221 HYKTONENGDNMNNKSLVTDLILGALQPSDVNKAIRKPSGLTAFESFGLGANSEKKEG 280
QY 1150 ID-----RVIEFISELYQONGKSNKSGTSDSAVIGTNTYENVGAYLHDIGNFD 1201
DB 281 EDDYDLVLCMLKNEKQIOECKKRRHGETSVENG--GKSCFTPLDTTLEE-----E 327
QY 1202 DCOSONE-----FCDEKS--DGKDNK-----YARDRKPDHDDAGCGSKSPTRVQ 1247
DB 328 PIBEENQVEARNICPKQTVEDKKEEETCTASPPPEKPVPH--VARRITTPPEVFK 385
QY 1248 IKTKKKAEEKDTECKTVNDILKENDGRQYEDCHPKKNSNGYDPDQGNINL-VEDEPRVC 1306
DB 386 I---WRGRNRKTKCEIYVAEMLKDKNGRTVGEVYRKET---YSEWTCDESKIKMGQHGAC 439
QY 1307 MPRRQKLCVHFLANDNEIKKLSQVN-LKEAPFKSAAPFPFSWY-YYSKSGEGNELD 1364
DB 440 IPRRQKLCVHFLANDNEIKKLSQVN-LKEAPFKSAAPFPFSWY-YYSKSGEGNELD 493
QY 1365 KELKEGKIPPAFLRSMFYTGDFGRFLFGTDI-----SKHGEGSKLKEQIDSLFKNGD 1418
DB 494 EKLKGLIIPDEFKROMFYTRADYRIDICLGTDISKDTSGVGC---KVKCNIDVF--- 546
QY 1419 QKSPNG-KTROEWWTESHSEIWEAMLCAL-----VKIGAKKDDFTENYGYNNVAF-S 1468
DB 547 YKISNSIRYKRSWENNGPIVMEGMLCALSYDSLNANNVNETIKKLEGNNEKEVIFGS 606
QY 1469 DKSTTLEERAKRPOFLWLTWETDYCYTRQYKYLKDOEKCKSKNDOLKCDTECNKR--C 1525
DB 607 DSSITLSKFSEERPOFLWLTWEMGENFECKEKKVYLACKPCOD--VDGSGKCKNGKCVAC 665
QY 1526 EDVVK-----YMKKKKEMIPQDK--YYKDRDKRRD--RQHGVAVMTDVTGT 1569
DB 666 KDQCKQYHSWIGIWDNYKKQKGYTEVKKIPLKKEKDVNSDADADYLTQDQNNKCV 725
QY 1570 NATDYLN-----RKFTASCGDRPGSASVYQRIQLLEKQAYYADADHCGCTFIENDK 1623
DB 726 NGTDDENCEYKCMHKTSSTSDMPES-----LDEKPEYKDK-CMC---VPRECN 771
QY 1624 YTNISS-----KDKCKGL 1636
DB 772 ALSVSGSGFPDGOAFGGGVLEGTCKGL 798

```

RESULT 5
US-10-153-273-4

```

Sequence 4 Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chilnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Willems, Thomas F.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE: ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
0-153-273-4
Query Match          3.2%; Score 617; DB 9; Length 1435;
Best Local Similarity 21.8%; Pred No. 1e-25;
Matches    332; Conservative 215; Mismatches 557; Indels 418; Gaps 69;
QY   1101 PACNENYNDITGLMKRQ---WNITSRKYLHQQAOMSVNSGSEASTAKNHIDRNVEE 1157
      K N I I I I I : : : : I : : : : I : : : : I : : : : I : : : : I
Db   20 KARNEY-DI-----KEKEKFLDYKKEKFNELDKKKYGNVOKT-----DKRIFF 62
QY   1158 L-SELYOONGSGSKNS----GTSDESAVTGTTTYENVAGATLHDGTGFDDCOSONEECDE 1212
      : : : | | | | | | | | | | : : : | | | | | | | | | | : : : |
Db   63 IEKNKDILNNKSFGRKRKRSYGTPDN---IDKNMSLIN-----                96
QY   1213 KSDGDNEKVAFRDPQDHDGA CGCKSGSKPRVQ--IKTKKAEEKOTEKVNDILKEN 1271
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   97 ----KHNNDEEMNNNYOSP.LSTSLLIKONKVPINNAVVRSLISPLDSR---INNCRNTS 149
QY   1272 DGKKOVEECHPKKNSNGPYDMQC GNINLYVEDRVCA PPRRQKI CVHFPLANDEIKKLQSQ 1331
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db   150 SNNEVLNSCRERRKG--MKWDCKRKN--DRSNYYCVIPDRRIDCIYNLS---IIKTYY 201
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY   1332 VNLEAFATKSAAAEFEFFSWYYYKSKDGEGNELDKELKEGIPIPAFLRSMEFYTPGVDRDL 1391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db	202	ETWKOHIETASKE---	SOLLKKNDDNNKYN-	----	KFCDLKNLSFLDYGHIA	246
Qy	1392	FGDIDSKGSEBSKLEQJDSLFKNGDOKSPNGT---	ROEMWTESHEITWAMLCALYK	1441		
Db	247	MGMDMEG--GYSTAKENKIOEYFKAGABEISEHKIKNFRKEMWNEFERKEAML-----	300			
Qy	1449	IGAKKDPEFENGYNNVFEJSDSTLJEFGAFKRPOLMUEWMDVOCYTPOKTKLOVOK	1501			
Db	301	-----SEHKNNIN--NCKNIPQEL-----	QITOWIKEMGEELFLEDRNSKLPKSK	345		
Qy	1509	CASNDOLK--CDTECKAKCKEDYVYK--MKKKWIPDQDYRYADREKFRORHIGVAVTDY	1566			
Db	346	CKNNLTLYEACEKECJDCPKMKRYDWIIRSEKFMHTLSKEYETOKVPE-----	392			
Qy	1567	TGNMADTYLNKRFETASGDKRGKSGASVYORNIQLEKAYYDADKHGCC--TKFIE----	1611			
Db	393	-----MAENIL--IKISNNKD-----	AKSLLLNNOCAEYSKYCDCKHTTLYSVLN	439		
Qy	1620	-ND---DKYTNISSKDKCK--GLVKEA--NTGALIKOMKCGNNNNNLKELTEDVLPFSRRL	1672			
Db	440	GNDNLTKEKREHJLDDPSKFGCDKNSVDNTKYMCKNNYILS-----TKDYCVPRRQ	494			
Qy	1673	RICFHALDGNNTIDPEYKDENL--RKRLEVAATEGYNLQYIKERKEKIKTSDAHKY	1730			
Db	495	ELICLNDIRIY-----DKNLMLKEHILAIAYESRLIKRKYKNDDKE-----	538			
Qy	1731	SYEPEPCSAMKSYFDLDDILIDGNLEDEKOTEENLKIFPKNGSTYSGKSDSTGTNP	1791			
Db	539	-----VKCIINKTFADIDJDIIGGDYMD--LSNRKVAKINTNSKIYHRRKNK----	585			
Qy	1791	GSTARKEFWNENKECWNAMICGYKGRGRDNGSGNSARSDDLKCGSVPSDDVPMGN	1856			
Db	586	DKLFDEDMWKVIKDYMNVAVLSWPK-----	DKTYCKE	617		
Qy	1851	RDEGATYOFLMFAEMGEDECEKHEKELKTVACANDYTOGDNEDKRKKCTACQYKFF	1910			
Db	618	DIENIIPQFFHFWSEMGDDYQODTKAIETLAKYCEKEKPEDDO-----NCKSKCSTYEW	672			
Qy	1911	ISEMFPYOEKOIKRYGE---NKDKIYSEHPAKDAEDAREYLDKOLKIKENSGDCEY	1966			
Db	673	ISKKEEYNKOAKOYOEYOKGNNNKMYSEKSIKPEYLLKTYSEK-----CSNLNFEDEF	727			
Qy	1967	K-----C--MKDYSQRLUDNGSQMMPASLDDEPREV--EGKCNCQVPRG	2007			
Db	728	KEELSHDYKKNCTMCPEYKDPIS--IIRNEQYOSQAVEEENTELAHRTLEPISSEPGK	786			
Qy	2008	PPVARETPEPSRVSLSKATASKEAKETAPTQPKVEVNTTEMBAQTRTBRAAQPRK	2067			
Db	787	NEOKREDDDS-----LSKISVS-----PENSRR-----	ETPAK	814		
Qy	2068	RSTATITTESYV--TMYKALLSNKPDPSRGISBCNPXTYGYPKMGCIYVSGKENENGIC	2126			
Db	815	DTSNLLKIKGVDSIMKRAVIGSSPN-----DNIN-----VTBEGDNISGV-	855			
Qy	2127	MPPRKKLCIINIQLYLYEFENKRDNDIEAFKCAIETQFLMKLYIENPAENELON	2186			
Db	856	-----NSKPLSDVAPDKKLELDONSESEYIV-----	NHLSKSPSINNIGDS	899		
Qy	2187	GTIDDEFRRIMTYTYGDYKDMFGTDISDKKIIITYVNSVTTILNENKKKODKKDEEL	2246			
Db	900	GS-----GSATVSESSSSNTGSLSIDDRNGDIFVFVOTANTEDVIYIKRENADKDE	951			
Qy	2247	RKIFEMKKNKFIWEGMILGYTLHJDBEN--EKKRIDNTOYNDMTLTPSLSEFPKRP	2302			
Db	952	KGADEER-----HSTISELSSPEEKMLTDNBSGNSLN-----HEEVEKH	990			
Qy	2303	QTLRMTETMAEEFCNKRKIDQLK-----LEAGCKEYEBNGSN-----	2339			
Db	991	TSNSDNYQSGGIVNNMVWERELKOTLENPSSSLDGBKAMHELSEPNLSSQDMSMNTGPL	1050			
Qy	2340	DGKIOEACAEACTYONFKIKMTYERQOREKKD-----KDGKTYKDPSTER	2388			
Db	1051	DNISEETTER--ISNNNEYVYNEBERLTLTKYEDIVLKSAMHNESSDGLGYLDE-----NS	1104			

QY 2622 GCANP1CDTWKYSFADL-----GDIY-----RGTDMLRIGGYLPVEILYKVE 2667
Db 1259 -----DLQKHGFHMMNNLHGCGVBERQ INSHHGNODRGN----- 1296
QY 2668 YIYKWKNNKKNKGRKYKDYVOTFSAMWDANKDIWAKMTKAPEDAKLF-RKGRMDGFER 2726
Db 1297 --SGNVLNMSNNNNNNNNI-----PSRYMLYDKKLDLDLXE- 1330
QY 2727 ITLQDKCGHDDPPVDYIYQFRFRWMTWSEYCKALME-----ELKFKKSCDH----- 2777
Db 1331 -----NRNDSTYKELI-----KTLAELNCCENETSVKCDHMTHEE 1366
QY 2778 --CSTSRCKNDIDENKCEQCKTRCOEYKKNFVLKWSLFDIOSNKY-----KELYEOPITY 2831
Db 1367 IPIKT---CTREKTRNLCCAVSDYCMYS-----FTYDSEYKNCYKREFDPSYT 1413

RESULT 7

US-09-924-154-17

Sequence 17, Application US/09924154

Patent No. US20020127241A1

GENERAL INFORMATION:

APPLICANT: Narmu, David L.

APPLICANT: Sim, Kim L.

TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use

FILE REFERENCE: 05213-0465 43170-262105

CURRENT APPLICATION NUMBER: US/09/924,154

PRIOR FILING DATE: 2001-08-07

PRIOR APPLICATION NUMBER: US 60/223,525

NUMBER OF SEQ ID NOS: 17

SOFTWARE: PatentIn version 3.1

SEQ ID NO 17

LENGTH: 1501

TYPE: PR

ORGANISM: Mammalian

US-09-924-154-17

Query Match

Best Local Similarity 3.1%; Score 594.5; DB 10; Length 1501;

Matches 31; Conservative 21; Mismatches 529; Indels 477; Gaps 69;

QY 2094 RGGIEGNCPEYTGQYKMCIGKSKENENGIOMPBRKLLINNOYL-NYETENKRN 2152
Db 125 RYIRGCKGR--STHTWIC-----ENKGNNTICIPDRVQLCITLQDKNSGETTDRK 178
QY 2153 DKEAFIKCALETQFLMLKYLIEPAENELQNGTIPDEFKRIYTYGYDKMEFFGT 2212
Db 179 LTRDKVFSAMETDILNKKYGRG-----FDDFCDDYKNSYLDYKDYIFGTD 226
QY 2213 ISNDKKIIVTSVTLINENKKKKDKKDELRK--IFEKNNKFTMEGHIYGLIYHL 2270
Db 227 L--DK-----NNISLVEES--LKRFEKKDSSVNLPTAMWRVYGRTRKMTIQPYA-HL 275
QY 2271 ---IDENEKERIRNYOYNDMTKLTPSLSEFVKRPOFLRPFTEVAEECH--KRRBOL 2324
Db 276 GCRKPDENE-----PQINRMLLENGKINCRMLKKEKEL 309
QY 2325 KLEAC--KEYCNGSNDGKTOEACAEVYTONFIKKWTEYEROREFKDKDGKY-- 2380
Db 310 TGECSVNRKKSDCSTG-----CNECYTYSLI-----NRQREYSILGKTYIK 353
QY 2381 -KDYSTERDIEKATCAHEYLNMKLELQGNKDCSCMOKPSQLPKTYQSOSSDAMP 2439
Db 354 VRYTIFRRIKVIQPNALDFELKNCSE-----C----- 381
QY 2440 ESLDYVPEEFNKCCEPCLSKSGSMITHKITEPKIPKNCVEKAAYLSKEAENNDITLK 2499
Db 382 KDIDKRP--FFEEYKGYEEK-----CMQOSYIDLKIQKRNDDICSFN 422
QY 2500 EKFTIESTKEKSKNSTNNPNCPKPKVYAPDKYIGRRNCEENRENFVDEYEMKCYK 2559

Db 423 AOTDIYSDDR-----CLEKKEKRP-----WKCDK 448
QY 2560 NSKFEYQEKRRVCPVPRRHHCLNLDI---KIERLDSNYLLKVRRTARRNGIDIIRN 2616
Db 449 NSEFYHHKGVCSPPROGCGNLNYLLNDIYVHNSQLLEI--MASKQEGKLMKK 507
QY 2617 ---FNSENGCANP1CDIMKY---SFADLGIYKTDMLRIGGYLPVEILYKVEYI 2669
Db 508 HGTIIDNANC-----KYNDSYDYKDIYIGNDLMDNNNSI-KYNNMLNIFERN 557
QY 2670 YGKWRNNKNGRNK-YNDVOTFSAMWDANKDIWAKMTKAPEDAKLFKGRMDGFER 2728
Db 558 FG---YKVGRRNKLFTIKELKNVWMLNRKNVWESMR-----GIDEVD 598
QY 2729 LQDKCGHDDPPVDYIYQFRFRWMTWSEYCKALMELEKFKKSCDHCKTSDCKNDY 2788
Db 599 QRRKTERIDEL--ENMOPFRFMSQWHPFCK-----EKEYWELKINDCKGNN 646
QY 2789 DENKCEQ--CKTRCOEYKKNFVLKWSLFDIOSNKYKELYEOPITYKISTYHVNQFVKL 2846
Db 647 GKSLODKTCQVNCINMNTWYTRKLAYELIOSVTKD-----DRKL 687
QY 2847 KTFKSECVSESEYELHETSK-CLNYKFNENDSSNIRTYAFETPKSYKACSC-----T 2901
Db 688 FSLANDKNVTF--LKEKAKKNSIDF-----TKIPIQDLKLFKRCSCMDYQ 734
QY 2902 LSKNPLNDPCPDQDKCKELQTFPCSKNDYDNNLDMNNAVYLVNSSDDNKGVLLPPR 2961
Db 735 LKVN-----KEMLSIDNSSEATDISEKGEBELVNNIN----- 769
QY 2962 RRLHCTRPITAYNKRCKEILKILKILTSASFQGLQOKYSEBELCEFAKMSYADYS 3021
Db 770 -----SVSVASGNKEI-----EKSKDEKOPKEKQ----- 795
QY 3022 DIKGTDMNDISLSEKIKIIFETSNEATENRKTWMENNRQIWMALGYYIATSKYTL 3081
Db 796 --TNGTLTVRRDKSDRRKNGKDTATD-TKNSP--ENIKVQ-----EHGTNGETIK 840
QY 3082 EGMQCPKDEETNOPLRWL-ITEMAQACKERKHVSDSLKTCPSNEDNPFASLBNQPG 3140
Db 841 EEPRLPSSSETLQSOBLEMAOKOKEEPPKKOEPEPPKKOEBKQOEBKQO 898
QY 3141 CONDRIKTIISLILIKNTMENILIKYKQDKOSSGNID-----NRPSEENV 3186
Db 899 -----EEBEOKOEEOQIODOSGLDQSSKVGAVASQNRNISSGQGNV 942
QY 3187 QSYIKSDOCALBLNDINE--IYTGKRNENNNEFEKYLKLYGLYFVEDETHKNVLD 3244
Db 943 KSSPEYVPOETSEBSSQDTKISTEPENNS--VDRATDSMLDPEKVINENMSD 998
QY 3245 GN-----IKKEEQTVR-----PKALY 3260
Db 999 PNTWTEPDASLAKDKKEVDKAKELQSVKRIEENEDVOSTPREDPTVYGKYGDAEM 1058
QY 3261 FFTPHVDSFYAPLPSHRAQYDPKNDILKSSLSVIVSAL--GLIALHFKMKKSKSS 3317
Db 1059 LFSPHATD-----NSESEGLNPTDDI-KTUDGVYKQEBIIGGGSATETSXNLKRP 1110
QY 3318 VDLRIILINIPQEGYGMPL-----ESKNRYIYPSGPKYKKTYYMGDPSGDEKTMMD 3372
Db 1111 KDY-----EPHEISEYVLSTGTTKESELSKSSITKTET-----DPRSNQOEATDD 1160
QY 3373 LSSSDITSSSESEYELDINDIYVGPSPRYTLLEVLEVP--SKR-----DIPSDDTPSNDT 3426
Db 1161 VVENSRRDNNLSNSVNOQSVNLNREDPLASFEVEVSEPPDSRIITTEVPSTTVKPPDE 1220
QY 3427 PRTNRFIDDEMNELKHPFYQYLPNT--EPNNNYKASADIPN--TEPNTLYSDNPEEK 3481
Db 1221 KRSEVEGEKAKETK--VAPVYPRALGEPMENSVSQSPNVEDVEKELISENN----- 1273
QY 3482 FIISIHRLDYTGKEISYINIMSTNTNNDIPMARNDSYGIDLINDS 3529
Db 1274 ---GLAH-DTHRG-----NISEKDLIDILH-LRNEA--GSTIILDD 1307

RESULT 8
US-09-924-154-14
Sequence 14, Application US/09924154
Patent No. US20020127241A1
GENERAL INFORMATION:
APPLICANT: Narm, David L.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentia version 3.1
SEQ ID NO 14
LENGTH: 1143
TYPE: PRT
ORGANISM: Mammalian
US-09-924-154-14

Query Match 3.08; Score 591.5; DB 10; Length 1143;
Best Local Similarity 20.88; Pred. No. 1.9e-24;
Matches 293; Conservative 188; Mismatches 453; Indels 473; Gaps 67;

QY 1118 NIIDKYEHEQOMSVSNGSIASSTAKNHNIDRVIEFLSELYQONGSKNSGTSDE 1177
DB 62 NIYSDVEDVANK-----LINSFENKSVK-----KRISFL-----NNTKTSID- 102
QY 1178 SAVIGNTTYEN-----VGAYLHDTGNFDDCOSONECD--EKSDKDNERYAFRDKPD 1230
DB 103 --IIPSYSDNFNSISENEDNSGN---TNSNPFANTSEISIGDKNQYFLQK-RT 155
QY 1231 HDAGCGKSGSKPTRVQIKTKKABEKDTECKTYNDLIKENDGKQVEDCPKNSNCP 1290
DB 156 HLFACGIR-----KSIKWICRENS----- 175
QY 1291 DMOCGINILVEDPRVCMPPROKLCVHFLANDNEIKKLOSQVNLKEAFIKSAAETPSW 1350
DB 176 -----EKIVCPYDPKRIQICIANFLN---SRIETEKREKELFISLVNTEA--KL 219
QY 1351 YYYSKDEGNELDKELKEGKIPAPLRSMPYTFGDRDLFGNDISKHSGSKLKEOI 1410
DB 220 LYNNK-----EGKDPSEIFCNELNRSFSDFNSEIGDDMDFG--GNTDRVKGXI 265
QY 1411 DSLF-----KNGDOKSPNGKTRQEMWTEHSHEIMEAWL-----CALVIRIGAKKD 1454
DB 266 NKRTSDYYKEKNVLEKN---NIKEMMEKKNANLNNHNVHKGNISRECAIIP----- 316
QY 1455 DFTENYGNVNVKFSDKSTLLEFAKRPQFLRMLEWYDDYCYTRQYLLKDVQECKSKNDQ 1514
DB 317 -----AEPQINLMIKEMNENFLMEKKRFLPLNKDCVEMK 353
QY 1515 LK-CDTECNKCEYVYMKKKKEMIP--QDKYYKDER--DKKRF----- 1554
DB 354 YEACFGGGRPLCSSYSTSPMKSKTOMEVLTNLVKKNSGVAKNFFLNDLFKRNKNNDLDD 413
QY 1555 -----DROHIGVMTDYGTNATDVLNKRFTASGDKPCASAVV--ORNIOLLEKQAYYAD 1609
DB 414 FFRNEKEVDLDCRCRYATITIKSLN-----GPAKDVUASQIVNVL----- 457
QY 1610 KHCCTFIENDDRYTNISKDKCKGLVKEANTGAIKQNKGNPNYNNLKELTEDVLEPS 1669
DB 458 RGFEC--NYSNNEKSWN-----CTG-----FTNKFPGTCE-----Pp 489
QY 1670 RLRLICE---HALDGNITDPEVKDENGILKRLMEVAATEGINLQYVKEKEKIKISD 1726
DB 490 RRQTLCLARTYLLHRGH-----EEDYKEHLGASTYEAOLLYKRYKEDENAL----- 537
QY 1727 AHKSYEVPSCAMKYSFYDLRIILIGIDNLED--EKQTEENLKKIFRNKNGTSVGKGS 1785
DB 1727 AHKSYEVPSCAMKYSFYDLRIILIGIDNLED--EKQTEENLKKIFRNKNGTSVGKGS 1785

DB 538 -----CSIIQNSYADLADIILKGSILIKDYKKMEENLKV--NKD---KKRNEE 582
QY 1786 TTGNPGSTARKEFWNNEKCYNNAMICGKRRGRDDGNSGARSDEDLCKGSPSDDY 1845
DB 583 SL-----KIFREKMDENKENYKMSAVLK-----NKECYKDYKFKIP----- 623
QY 1846 PMGNRDEGTAYOFLRWEAWEDEFC--KHKEK--ELEKLVACNDYTCGDNEKRRKCTD 1902
DB 624 -----OFLRWEKEMGDDFCERKEKIKYSFSEFYECKKKDCDEN-----TKN 666
QY 1903 ACTQKRFISMKQYEQEIKKYGEND--KIYSEPAKADAREYIDKQLK----- 1954
DB 667 KCSYKRWIDLKSEYEQVDYTKDKKKMYDNIDEVKNE--ANYLAKESKCKCKDVNF 725
QY 1955 --KICENKSGCEYKCMK-----DVSTQRLDGSQNNPASP 1990
DB 726 DDKIFNESPNYEDMKCKKDEIKYLNELKPKTKHDIYDIDTFSDTREDGPIISIMANIN 785
QY 1991 DEP--KEYEGKCNQVPRGPRVRRETPSPVSLISKATA--SKREAKTAPPTQPKKVE 2046
DB 786 EQQSGKDTSNNGSETSDSP--VSEHPESDAIINVEKLSGDESSSETRGDIIDNP--SVT 842
QY 2047 NLITEMRQOTRRRAAOTRRTS-----TATTESD--VGMVYAILSNKPDNRGI 2097
DB 843 NNVEVHDASNTQGSVSNFSDITNGHSESSLNRTTNAODIKIGSGNOSDNQENSHSS 902
QY 2098 EGCNPKTYGYPRKGCICVKSRENENGCMPRRKKLCLINNIQYLYNETENKRRNDIKEA 2157
DB 903 DNGSLITIGVPR-----SEDN-----TQNTYIS----- 925
QY 2158 FIKCALETOFLMLKYLIEENPAENELONGTIPDEFKRMVYGYDKMFGGDISNDK 2217
DB 926 -----QNPBRDTPNALASLPD-----DKINEIEGFDSSRDS 957
QY 2218 K-----ITTVNSV--TILLENKKKKODKKDELRKIFWKKKFTWEMITGLTH- 2269
DB 958 ENGSDTTSNTHDVARTIVSERVNSHD-----FTRNMANNNAHQ 1000
QY 2270 -LTD--EN-----EKEKIRNDYQYNDMTKLTPSLEEFYKRPQFLRMFEAEEFCNKRKE 2321
DB 1001 YITQIENNGIIRGOEESGNSVNTKD-----NPKSNPSEENDHKKNOIEVSRQTKRARE 1056
QY 2322 QLKLEACKEEYECNG-----SNDGTQBC--AEACVYQNEFK 2358
DB 1057 EIKIKS---KONKNNEYSMECYTSDERNSSPGCSPEERKKILCCQISDYCLAYNF-- 1111
QY 2359 KMKTEYERQREKFKKDKGKRYKDYPS 2385
DB 1112 -YSIEY---YNCISKEIKSPYKCFKS 1134

RESULT 9
US-10-153-273-6
Sequence 6, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:

Query Match.	2.28;	Score 431;	DB 9;	Length 749;
Best Local Similarity	21.5%;	Pred. No. 7.7e-16;		
Matches 210;	Conservative 128;	Mismatches 262;	Indels 378;	Gaps 44;

[illegible]

```

RESULT 10
US-09-924-154-16
: Sequence 16, Application US/09924154
: Patent No. US20020127241A1
: GENERAL INFORMATION:
: APPLICANT: Narum, David L.
: APPLICANT: Slim, Kim L.
: TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
: FILE REFERENCE: 05213-0465 43170-662105
: CURRENT APPLICATION NUMBER: US/09/924,154
: PRIOR FILING DATE: 2001-08-07
: PRIOR APPLICATION NUMBER: US 60/223,525
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 16
: LENGTH: 972
: TYPE: prt
: ORGANISM: Mammalian
: US-09-924-154-16

```

Query match	2.0%;	Score 382.5;	DB 10;	Length 972;
Best Local Similarity	19.7%;	Pred. No. 5.2e-13;		
Matches 253;	Conservative 157;	Mismatched 450;	Indels 427;	Gaps 49

QY	1051	LRMTLEAEWYCVQKCEKDKLKECK-BCK-----DKDNQGGCTKESGCTCTCTEAC	1107
Db	16	INNM-----LKIINIVTEYQKKEICKSCKCKKAMIDKNND-----FTLLSEIY	60
QY	1104	NEYNDITGLKEQWNTIS--DKYELHBOAOMSVSNGSIGEASSTAKNHHIDRNVLEFLSE	1160
Db	61	LKKNKSSLYLKYTFEYIKQKMPDKYELN-----FSSI	92
QY	1161	LYQNGSKNSKSGTSDBSAVIGTNTITYEWGAYLHDTGNFDQSQONEFCDEKSDGKUNE	1220
Db	93	FDDLANKRYNK-----CIQNNKRIENNALYV-----KIEDICN-----	125
QY	1221	KYAFRDPQDHDGACGCKSGSKPFRVQITFKKAEKDECKTVANDILKENDGKQYVDC	1280
Db	126	-----NTKKSITLYGELYCKE-----KGN-----	144
QY	1281	HKKSNNGTIPDMOCGNINLVEDPVCMPBRKOLCVHFLAND-----NEIKKLSQYVNLK	1335

```

Db 145 -----KIWOCINEHIDPEVCGPPRRQOLCGNLDKDEKFNVDKFLNEL---192
QY 1336 EAFIKSAAETFFSWYKSSKDEGNELDEKELGKIPAFILBSMFYFDYDFLEGTD 1395
Db 193 ----ILGIRDSGKFLIEKRYKRNHMHNYLDERA-----CKYANSFDDYKNIILIGKD 240
QY 1396 -----ISKHGESKLUKEQIDSLFKNGDOKSPNGKTRQEWTEHSHEIWEAML 1443
Db 241 MWBDPNSIKTENILKGFEEB--IKANIVSMYPSYADLSL-EPFKHMDONKQJLMEALS 297
QY 1444 CALVKIGAKKDDPLENGVNNVNFSDKSTLEEFARKRQPLRMTEWYDYCYTRQKYLK 1503
Db 298 C-----EFKGNHT-----GVCLEMDNDNDYLMHREKMKPCJDKLKWMD 339
QY 1504 DVOEC-----KSNPOLKCDTECNKCKEDYVYK-MKKKEMIPQDKYKDERDKR 1553
Db 340 VIKPCIDKKVKSPKSENSDVATVONKSCJTDYDKWILKKRKKYKQSSKRYK--RDRSL 397
QY 1554 PDR--QHIGVMTDYGTATNATDYLNKREFTVASCQDKPGSASVQORNIQLERQAYYADKH 1611
Db 398 FNNVIOQL-----KPEYLSMKCT-----E 417
QY 1612 CGCTKFIENDKXYNI-----SSKDKCKGLVKEANTGAIKQNKGNPNYNNLKELEDFYL 1667
Db 418 CTC-----NLDTOFYYPYKGYEDICKSYKp-----444
QY 1668 PSRLRLICFHALDGNATDPEKNGLRKRLMEVAATEGYNLQYKKEKKEKIKTSDA 1727
Db 445 -----YDPEIKDE-----EF 455
QY 1728 HKSYEYVPCGSAMKYSFYDLRDIILGIDNEDEKOTEENKIKFNKNGT-----SVKG 1781
Db 456 NEPSLANPIS--LTSQDTERSSVDVLSIKENVD--LKPEFKPGKQSSHVQVGN 510
QY 1782 GSSTJGNPSTARKFVWNE-----NKECVNANMICYKRRDQNS-----GNSARSDLD 1833
Db 511 PRESSESPSANGREDBSTSSSTYNDGVITSSSLGSSSDRVSSSPVGVDEHEAHEL 570
QY 1834 ---KKGVSDDDYPMKRNDEGTAYOFLRWFALWGEDECKHKEKELEKLVGACNDYTC 1890
Db 571 PPKIIDGVQSDBESTLSQKGESSQEQ---HNLDGSSLSRHSNDEERSI--ITSUVEH 625
QY 1891 GDNEDRKKCTDACTQYKKTISEKPOYEKQIKYKGNKIKYSEHPVAADAEADAREYLD 1950
Db 626 GTNSLFGSOIQDET---ILGESEPLTSP---PEHETSKMDTHAGKRMGEVRRASY 677
QY 1951 KOLKIKCENKSGOEYKCMKDVSTQRLTDGNSQMPAS-----LDDEPKVEGKQNC 2002
Db 678 DSSSEMSNGRGGLKTKEMGEVYTGITSKNDINLEDSTVHSRONKLENSGNDTQGEHI 737
QY 2003 QVRGPPRVRRRETPSPRVLISKATASKKEAKTAPTQPKYKENTLTTEKRAOTRTERRA 2062
Db 738 NVLQ-----MDKHLNPPISERDSV--LESEFSKLNRT-----770
QY 2063 QQRRTKRTSTATTESDVGIVAKAILSKRPPDSRGIEGCPNKRYGQYKWCQIV-----2115
Db 771 SHTHDNRRITETIENNIGGLSINSVHDGRDSORRMRHINSRHSRHSLESIDIVAGDISN 830
QY 2116 ---GKSKENENGICMPR---RKLCINNOIYVANYTEKKRONDIEAF--IKCALIET 2166
Db 831 IEGGEEEDANTLKYPRNVLNNKNSRTYNIIEYIVDVKNVADDIMRSYKSNRC---T 886
QY 2167 QFLMKLIENPAENELQNGTIPDEFKRLMYYYGYDKMFFG-----TDISN 2215
Db 887 NNLSSNVCCK--LKRSLSLNTCTNEDSKRLCSISDYCMKFFNFNSGTHSCMRKEFSN 943
QY 2216 DKTIITVSVTITLNNKKKODKK 2242
Db 944 HAYKCFAGKGFSSMFLDKKKKKKK 970

```

```

US-10-153-273-20
; Sequence 20, Application US/10153273
; Patent No. US20020169305A1
; GENERAL INFORMATION:
; APPLICANT: SIM, Kim L.
; Chitnis, Chetan
; Miller, Louis H.
; Peterson, David S.
; Su, Xin-zhaun
; Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; AND PLASMODIUM FALCIPARUM ENTYROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knodde Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0; Version #1.25
; APPLICATION DATA:
; APPLICATION NUMBER: US/10/153,273
; FILING DATE: 21-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 411 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: Internal
; ORIGINAL SOURCE:
; SEQUENCE DESCRIPTION: SEQ ID NO: 20:
US-10-153-273-20
Query Match 1.98; Score 365; DB 9; Length 411;
Best Local Similarity 31.88; Pred. No. 1,5e-12;
Matches 84; Conservative 17; Mismatches 137; Indels 26; Gaps 5;
QY 858 VIMPPRRRHICTSLNLEHQTDDHPLNGNIYVDLYVNSFLGVDLLSAKYENKRIIRYK 917
Db 51 VFLPPEHHCSTSLNXXXXX-----XXXXXXXXXXXXXXXXXXXXXXXXXXXXX 104
QY 918 NNLKGPREVDPKHQTITICRAIRYSPADIGDIIIRGLMERNGDMVRLQGLFVPGNIH 977
Db 105 XXXXXXXXXXXXXXXXXXXXAMCAVRYSPADIGDIIIRGDMDEKSSXXXXXXXXXXXXX 164
QY 978 KSLKGRKNDKYND-----EYKLRLRMWMEANAKYWEAMKCDIKLKLKSGHQSTQSS 1093
Db 165 XXXXXXXXXXXXXXXXXXXXKPKLRADWWEANRQVWRMKC-----ATYXX 212
QY 1034 YCGYSDHTPLDDYIPKLRMTWEAEWYCKYQKREYDKLEKCKECKDKNDGCGCTRESG 1093
Db 213 XCXXXXXX-----XKIPQLRMWTEWAEWYCKAQSQYDKLKLKXXXXXXXXXXXXXCK-XX 268

```

OY 1094 TGCTKCTACNENYNDIIGLMEQM 1117
 Db 269 XXGCKKACADCKYKEETEKWNEQM .292

RESULT 12
 US-10-153-273-2
 / Sequence 2, Application US/10153273
 / Patent No. US20020169305A1
 GENERAL INFORMATION:
 APPLICANT: Slim, Kim L.
 Chitnis, Chetan
 Miller, Louis H.
 Peterson, David S.
 Su, Xin-zhaun
 Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660

COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/153,273
 FILING DATE: 21-May-2002
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/210,288
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1115 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Protein
 HYPOTHEICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium vivax
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-10-153-273-2

Query Match 1.88, Score 343.5, DB 9, Length 1115;
 Best Local Similarity 18.88, Pred. No. 8.9e-11;
 Matches 243; Conservative 177; Mismatches 435; Indels 439; Gaps 60;

OY 2304 FLRWTTWAEECRNKKRDKLLKLEAGCKEYEGNSNDKQTQECACACTYONFIKKWTE 2363
 Db 58 FLHHSQVNNVILLETITLIE---CNEYVYKGGNGYKLLAK-GHHVEEDN-LEK--- 108

OY 2364 YERQREKRRKDKGRKYDYSTERIDIKATCAHEYLMMKLE-----LCGNK----- 2411
 Db 109 --LQGTNRSRSENIKYK-YGVTELKITYAQMGKRSRIILKESIIYGANHGNSYMGK 165

OY 2412 ---DSCNQXSSQLPKTQOSSDANDMESLDYVEEFNKCCEPLSKKSGSIHTKK 2468
 Db 166 DGGDTGEGKEGH--KTDSTDNKGANLVMIDY-----ETSSNGQ----- 206

OY 2469 ITPFKIPMNCVEKAAAYLSKEAENNDITLKEKFIPIESTKREKSNSTNNP--CDPK 2526
 Db 207 -----PAGTLNDVLEFVYGHGNS-----RNSNGCNPYDIDHK 241

OY 2527 KPYAP---DKYIGRRNRCNRENEKVDYEMKCYKNSKFOEKKRVCPPREIMCLRN 2583
 Db 242 KISSAIIHAFLONTVMKNCYKRRRRRDMDC-----NTKDDVCIPDRYOLCKE 294

OY 2584 LDEIKTERLKDSN-----YLKMYRTARNEGDIITNFENSEGCANP1CDTMR 2633
 Db 295 L--TNLVNNTDNTNFRDITFRKLYLKRKLIDAAVEG--DLILKLN--NYRYKRCCKDIR 349

OY 2634 YSFADGDIRGDMRLIGLYLPVEIKLYKVEEIIYGMKRNKNGKRYNDVQTFRSW 2693
 Db 350 WSLGDFGDIIMGTDMGI--GYSKVENNRISLF-----GTDE--KAQORRQW 394

OY 2694 WDANRDIWKAMTCRAPEDAKLFRKGRMDGFEKITLIDKCGHKDDPPVDYI----- 2746
 Db 395 WNSKAQIWTAMVSY-----KKRLKG-----NFIWICKLVN 426

OY 2747 -----PQFRMWTSEYICCALMELEKFKSCDCHCTSCKNDYDENK--CE--QCKT 2798
 Db 427 AVNIEPQIYRMIREGRDYSELPTVOYKLEKCDG-----KINYDKKCKVPCQN 479

OY 2799 RCOEYKNEFLKWSLFDIOSNKKELYOPITYKISTYDHVQNFQKLTFRKSESVESF 2858
 Db 480 ACKSDQMTTRKKNQMDVLSNKF-----ISVKAENVQIAGIYTPYDILKOE--LDEF 530

OY 2859 SEYLHETSKCLNFKYKENDGSSNIRTYAFEEPKSYKACSGTLPKSNLDCNPTQNKD 2918
 Db 531 NEVAEN-----EINKRDG-----AYELCYCSV-----EAKK 559

OY 2919 GCKELOTFTFCSKNDYNDNNLWMNAVLYLNS-----DDKGVYLIPIRRRHLCRPIITA 2972
 Db 560 NQGEVYT-----NVDNAKSQATNSNPISQPYDSSKAERVDPDSTH----- 600

OY 2973 YNRRKDEILKLLKLLTSAFSOGQLLGOKYKSEELCFEAMKYSAVDSIDITKTDMDT 3032
 Db 601 GNVNGQDSSTTGKAVT---GGQ--NGNQTPAESV-----QSDIAESVSAKNV 646

OY 3033 SLSEIKKIFETSNATNRTNRTWENNRRQIWMALGCIKATSKVTLDEGWCQLPRDEE 3092
 Db 647 DPQKSVK-----RSD 658

OY 3093 TNOFLRWLLEWAKOCKEKKHVSLSLTKCPRSNEDNFEASBLRQCCONDIKYISLN 3152
 Db 659 TAS-----VTGLAEAGKKNLGAANS-----RPSSTVEAN---SPG--DDTVNSASIP 701

OY 3153 ILIKNTMEN-LNIRYKQLKQSSGNIDNKPSEE--NVOSTYKSKDSQCALNDINEITY 3209
 Db 702 VV---SGENPLVTPYNGLRHSKSDSDGPAESMANPDNSKGETGK-----GQDNMDAK 753

OY 3210 GTKNNEENFEKVLKLLPGLYFEDETHKNHVLIDNKIKEEQIVRPKALYFTPHVDSF 3269
 Db 754 ATKDSSNS-----DGTSSATGDT-----TDAVD-- 777

OY 3270 YQAPLFTFRVAQYDPKNDILKSSISVIVSALGLIALHFKKKFKSSVDLRLINIPG 3329
 Db 778 -----REIN----- 781

OY 3330 EYGAFTLESKRRYIPYSGPYKGTIYMEGDISGDED-----KYMDLSSSDITSS 3382
 Db 782 -KGVP--EDRDKTVGSKDGGGE-----DNSANKDAATVVGEDRIRENSAGSINDR 829

OY 3383 SEYEELDINDIYVPGSKRYKTLIEVLEPSKRDLPISDITSNPPR--NRFDIDENEL 3440
 Db 830 SK-NDTEKNAGSTDSQSEDAITALSTESLESTESGDRFTDNTNLEKNGEKERDQ 888

OY 3441 KHDFVSOYLPNTEPNNNYKSAADIPNTEPNTLYSDNDEEKPFIISIDRDLTYGKEISYN 3500
 Db 889 KHDFKSDTPEEENSD-QTTDAE-GHDRDSIKNDKAEKRRKM-----NKDTFTKNTNSHH 942

QY 3501 INNSTNTNDIPNANRDSYRGIDLINDSLVYN 3534
 Db 943 LNSNNLSNG-KIDIKERYKRDYKATREDIILMS 975

RESULT 13
 US-09-815-242-12996
 Sequence 12996, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 12996
 LENGTH: 6281
 TYPE: PRF
 ORGANISM: Staphylococcus aureus
 US-09-815-242-12996

Query Match 1.5%; Score 291; DB 10; Length 6281;
 Best Local Similarity 17.4%; Pred. No. 7.4e-07;
 Matches 713; Conservative 535; Mismatches 1587; Indels 1260; Gaps 173;

11 KGNMAASLEGDAKSPITKESHS-----ARNVLEKYAKNIRHPSKYAKEHYDSLKGL 64
 Db 1974 QVNSASALMGDEKLAQAKOTAKSDIGRLIDLNNAOQTANAEVDAQPNLAATVAAKNKA 2033
 QY 65 TRKEFGGSPYVKNKNIYYPCNDLH-----KEHTNLRKDDVNLKRGHGRONFEDD 120
 Db 2034 TSLNTAMG-----NLKHALAEKDNK-----RSVNTDAD 2063
 QY 121 ESEECGNKIRYKRNDAIACAPRRRHMDKN-----LEALINDITQNIHDLG- 170
 Db 2064 QP-----KQAAVYTAATQAEAITNANGSNANETQVQALNQLN-QAKNLDNG 2110
 QY 171 NVLVTAKYEGESTIVNNPHKGTSDACTALARSFADIGDIYRGIDMEKPNVHDYETGLRE 230
 Db 2111 NKVAOKESAKRALAS--YSNLNNAOSTATSOIDNATVAGVYAAQ-NTANELNTAMGQ 2167
 QY 231 VFKEHIDGMEDEYKNYN-----PDGSGNYY-----KLEAWNNVNRK 269
 Db 2168 LONGIND--ONTYKQOYNTFDADQKKDDATTAATNAAGITLDRAGHGNMTKAOVEALNQ 2225
 QY 270 VW---EAITCDASYKSGYFMOSESNTPLFSNPACGHKQ-----GKVPNT----- 310
 Db 2226 VTTAKNKLNDANDANVROAK-SDAKANLGTLLHNNAAQKODLSIQEGATTYNGVNGVYKTKA 2284

QY 311 --LDVYPQILRMFDEWGEFRCRRNITKLKVDKSCNDKRELYCSINGDCTTTTKKGI 368
 Db 2285 QDLDGAMORLQ-----SALANKDQTKASBNYIDADPT--KK-- 2318
 QY 369 LHLDNKCTDQSTFKCKVFEWVLGNQOEAFFKQKKEKYEKEIOSYLSNKRFPVNNINSEYQO 428
 Db 2319 TAFDNIITQA-----ESTL-NKDHGANKDQAOVDAIOSYSTENALNGDNLNLO----- 2366
 QY 429 FFEKLKETQATNDPFLNL-----LNEGKYCKGGLPGEKDIPTFSNADKGIYRS 479
 Db 2367 -RAKTEAIDAIDLNLTLNTPQKFTALKQOVNAQVSGVTDLKNASATLNA----- 2416
 QY 480 EYQVCPDQGVKCDGIKTYHKSNDNREVRNNEYKPPKGVKPTNITYLYSGNEGDIYOK 539
 Db 2417 -----MDLKOALADHD-----TIYASG----- 2434
 QY 540 LENECSSTNYKDKNNQKMECYKKDINIRCKLEONTFLINDPKIISHNFEFLAVTYL 599
 Db 2435 --NTYNAS--PDKQAVYDAYNAKN-----IYNGSPVIT--NAADVTA-- 2474
 QY 600 LBDTIKMNDKLKTICINNTTTHCIDECNRNCLCPDRWVVKQKEEWSIKKLTFRKKNIQOS 659
 Db 2475 -----IQRVNNAETGLNGDIN-----LATAKQAKDALRQMTLHSDAQKOS 2515
 QY 660 YSNINNLFEGYFFKYMDKLDKDEAKWKELMENIKRKKNEFSNLENNRDYLENAIELLD 719
 Db 2516 ITGDISATQ-----YTGQVSKDNATMLDAMMOLRNSI--AND 2554
 QY 720 HKET-----ATICKNNNTNEA--CETSHNATNTNCPVPRGSTOPTKNKELAQFFKS 771
 Db 2555 DVNASOPYDADRDKONANITAVTNAENITNATSOPTLDSAVTAQANOVST-----NKT 2609
 QY 772 AVEEARNRGLHLKRAHGEIYKRGRRKDKNLCRIMIKHSN-----NLGFSNGCP 825
 Db 2610 ALNGAQNLANKKQETANIN-----QSLHNNNQKODLNQVYNAP- 2650
 QY 826 DKGSTGDIOT-----REVVGTEWVDEPHMKKHEDVIYMPRRRHITSNLEHL 875
 Db 2651 -NISTVNOVYTKAEOLDQAMERLINGIQ--DKDQKQSVNFTDADPEKQAVYNAVTA 2706
 QY 876 QTDHPLNGNITVDLVNNSFLGDVLLSAKEYEANKIIRMYEKKNNLKGPKVETPKHOT-- 933
 Db 2707 ENIIINOANGTNANQSOVEALSTVYTT-----KALNGDKRYVTDKANNANQ 2752
 QY 934 TICRAIYSEAD-----IGDIIRGRDLMERNGDVVKLGHLFEGNIHSLKCK-----GN 985
 Db 2753 TLTSLDNLNNAQGAAYTGNINQAHYAEVY-QAIOGAQELNTAMGNLKSNDKDTTIGS 2811
 QY 986 DKYNDAPEKYLKLENNWE-----ANRAKVEAMKCDIK 1019
 Db 2812 QNFADADPE-----KKNAYNEAVHNAENILNKSTGTVPKQDVEAANOVVATKAALNGTO 2867
 QY 1020 YLKDKSGHOSTOSSYGYSHPFLD--DYIPQKRWMTWMAEWKCYOKKE-----YDKLK 1073
 Db 2868 NLEKAKOHANT--AIDGLSLTNAQKEALKQLOOSTYVAEAOGNEDKANNVDAAMDKLR 2925
 QY 1074 EKCECKDKDNGOGCTKESGTCTKEACNEBYNDIILGLEKQENIISDYKELHDA-- 1131
 Db 2926 QSLAD-----NATTKQN-----QNTDASQNNKKAQVYNAVYTAQOIIIDQTS 2967
 QY 1132 ---QMSYNSGIEASSTAKHIDRNVIEFLSELYQONGSKSKSGTSEASVIGTMYE 1188
 Db 2968 PTLDPYVINOAGOVSTTKKALNGN--ENLEAAKQO--ASOSIGSLDNLNNAOKOTVTD 3022
 QY 1189 NV-GAYLHDTGNFDQSONEFCEKSDGDKDNKRYAFROKPOHDGACGCKSGSKPRTVQ 1247
 Db 3023 QINGAHIVDEANOIKONAQN-----LNTAMGNLKQALADK-----DATKAT 3063
 QY 1248 IKTRKRAEKEDECTKY-----NDIILEKNDK-----KQVEDCHPKKNSNGYPDW 1292
 Db 3064 VNTFDADQAKQOAYNPTAVTNAENIISKANGNMTOAEVEGAIRIQVNAAKQALNGN----- 3118
 QY 1293 QCGNINLVEPDRVCPMPRRKCLCVHFLANDNETIKLOSOVNLKEAFLTKSAAEFTFSWY 1352

Db 3119 --ANVOHAKBEPAL-----INSSNDINOAKDA-LKQ--QVONATTVAGVNN 3161
 QY 1353 YKSDGBGNELDEKELKGIIPPAFLRSMFYTFGDRYDFLGTDLISGHGSGSLKQIDS 1412
 Db 3162 VKQTAQELNNAHQOLKQ-----GJADEQYTKADG 3190
 QY 1413 LFKNGDGKSPNGKTRQEMWTEHSHEIWEA-----MICALVKIGAKKDF 1456
 Db 3191 NFNAPBDKONA-----YNOAVAKABALISATPDVYVPPSELTALNKTQAKNDL 3241
 QY 1457 TENYGV-----NMYKFS-DKSTTLEEFARQFRLMTE-----WYDDCYTRQKYK 1503
 Db 3242 NGNTNLTAKQONVOHAIIDOLPLNQ--AQREYSQITQATVYVNAV100AATLMDAMT 3300
 QY 1504 DVOEKSDNDOLK-----CDTECNKCEDYKVMKKKKEMWIPQDKYKDEKDKKPR 1556
 Db 3301 QLKQJIANKQIKGSENYHADTD--KQATVNAVTKABELKO-----TNTPTMDP 3350
 QY 1557 QHIGVAVTYTGYN-----ATDYN---RKFTASCGDKPGSAS 1591
 Db 3351 NTLQALTRVNDTNQALNGNOKLADAKODAKTTLTGLDHLNDQKQALPTQVQAPDIAT 3410
 QY 1592 V--VORNIOLLE-----KOAYYDADKHGCTFEIEND---DKYTNISSKDKCKLVE 1639
 Db 3411 VNNVKNQONLNAMTNLNNALDQKTETLNSINFEDADAKKDAYTNAYS--HAEIILSK 3468
 QY 1640 ANTGAIRKONKGPNNYNNIKELTEDVLFPSSRLICFHALDGNVYDPEVKDEGLKRLM 1699
 Db 3469 AN-----GSMASQT--EVEQAMQRYNEAKQALNGNDNQYRAKD-----AA 3506
 QY 1700 EVAATEGYVLQYKFE--KKEKEKITSNAHKYSTEVPSCAMKSYFDLRIIGIDL 1757
 Db 3507 KOVITNANLNOAQKALQOVDAAQT-----VANVNTIKQADOLNAMOQLKQ 3557
 QY 1758 EDEKQTEENLKKI-----FNKNGSYGKSGSDSTGNGPSARKFEMENKECWNAM- 1810
 Db 3558 IADKQDTKANGFNVANDDTKQANAYNAVAHAQIISGTNA-----NVDQOYVAQALQ 3610
 QY 1811 ICGYKRGDDGSGNSARSDEDLKKGVSPPDDYPMGRNDEGTAYOFLRMFAE--MGE 1868
 Db 3611 QVNOAKGDLNGHNLQVARDNANTALDLPN-----LNQOKTALQOVSHAEVLGV 3663
 QY 1869 DECKHKEKELTVGACN-----DYTCGDNEDEKCKCTDQYKKTISEK 1915
 Db 3664 NAIKQNALNNAAGTLKQOIOANSOPVSDPTQAD--ODKQOAVNMANAQO--IANGI 3721
 QY 1916 P-----QYEQKIRKYGKNDKIYSEHPVAKADADAREYLD-----KOLKIKGKSG 1962
 Db 3722 PPTVLTPTVTOAVTMMQAKDLNGDEKLQAKQELANLDTLMDLNOQORDALRNQIN 3781
 QY 1963 DCEYKCMKDVSTQRLTDGNSQMPASLDEKREVEGKCNCOVPRGPPRYRRETPSPRVL 2022
 Db 3782 QAO-----ALAVEQOTKQNAQVNTAMSNLKGINKDTVKASENYHADADKQATYNA 3836
 QY 2023 ISKATASKKEAKTAPTRKOPKKEVMTJTEEMRQOTRRAAQOTRKRRTSTATTESDVGM 2082
 Db 3837 VSQAGIINO--TTMPTLNPDEITALTQV--TDKNGLNGEALAVEKONAKAAYVSGM 3891
 QY 2083 VKAILSNPDSRGIEGNCPTGYGPRMGCIKSGKENENGICMPPRKKILCINNI--- 2139
 Db 3892 THLNDQAKQALKGQID-----QSPET--ATVNGVKTQATNSLDQAMQLOALNDRKOT 3942
 QY 2140 -----QYLVETENKRDNDIKFAFKCALIEFOFLMKYTIENPAENELONGTTPDEKFR 2195
 Db 3943 LADGNVLAADPD--KONAYKQAVAKAE-----LNLKQSGTNEVQ----- 3980
 QY 2196 IMYVYGDYKMFCTDISNDKTIITVNSVTIINENKKOKDKKDEELKRTKFEWENK 2255
 Db 3981 -----AOVESTNEVNAKQALNGNDN-----ANA 4006
 QY 2256 KFIWEGMIGLYLHJTD--ENEKIKRDNYQYNDMTKLTPSLEEFYKRPQFLWETEMA 2312

Db 4007 KOQAKQOLANLT--HLNDQAKQSESOITQAPLVTDVYTTINQAKQTLDAHAMELR----- 4059
 QY 2313 EEFCKRKQOLKLEAGCKEYCNOSNDKTOECACVYTONFKMKTEYER----- 2366
 Db 4060 -----NSVADNOUJLASE--DYHDATQORQNDYQAVTAAN 4093
 QY 2367 -----OREKFKKDKGKTKKDYPSERDLEKKTACHIEYINMLK 2405
 Db 4094 NIINQSTSTPMNPDVNGATTOYNNITKVALDSD-----ENLAAQOANNRLD 4141
 QY 2406 ELGKNKDCSCMQKPSQOLPKTQOOSQSSDANMPESLDVYPEEFNK-----CECPELSCK 2460
 Db 4142 OL--DHILNNAKQOOLQOI--TQSSDIAVNGHQKTASLNTAMGNLNLATADHQAVER 4197
 QY 2461 GSMITHKTKTEKIPMN--CYEKAATYLSKEANNDITLKEKFT----- 2503
 Db 4198 GNFITAND--TDKQATVYNTAVNEAAMINKQOGQANQREVEQALTKVOTTLQALNGDHL 4255
 QY 2504 PIESTKEKSKSWTNNNPCDEPKPYAPD----- 2532
 Db 4256 QVAKTNATQALDALSLN--DPQKTLADQYTAATLVTAHQIEQANNTLQAMHGLRQS 4313
 QY 2533 -----KYGRNPNCEENRENRK-----VDYEMKCYKSKFYQEKRRVCVPRR 2576
 Db 4314 IQDNATKANSKYINEDPEQOQNYDOAVOANNITINEQTATLDNNAIQAAATVNTTKAA 4373
 QY 2577 EHMCLRNDEKIERLKD-----SNLTKMVRTARNEGIDIKN 2616
 Db 4374 LH-----GDVKLQNDKQKHAQVQSALHLNNAOKHMEDTLIDSETTETAVAKODL----- 4422
 QY 2617 FNSENGCMNPICDPMKSFADLGIYAGTDLIRIGLYLPEVEIKLYVFEYIYGKMRNK 2676
 Db 4423 --TEQAQLDQLMALQOOSIAD--KDATRASS-----AYVAE--P 4456
 QY 2677 NKGRKRYNDVOTFNSAWDANRDKIWKAMTKADAKLEFRGRMDFERITLLQDK-- 2733
 Db 4457 NKQSYDEAVONASISLIGLNNPTINKGNVSSAQ--AVISSKNAIDGVER--LAQDKOTA 4513
 QY 2734 --CGHKDD--PPYDVLPOREFRMWTESEYK--ALMELEKFKKSGCDKTSR 2783
 Db 4514 GNSLNHLDLTPAQQALENOINNATTRGEVAOKLTQALQNMAMEALRNSIQOQOOTE 4573
 QY 2784 CKNDYENKCEQCTRCQYKFNVLKMKSLFDIOSNKKELYEOPYTKISTYDHYONFV 2843
 Db 4574 GSKFINEKDPQ-----KDAYQAVQNAKDLING--TNPPTLDKAQ--V 4612
 QY 2844 OKLTFFSECSVESEFSEYLETSCILNYK-----FENQSSNIRKYAE----- 2888
 Db 4613 EOLIQ-----AVNOAKDNLHGDQKLADDKQHAVTDLNQLNGLNPORQALESOINNATR 4667
 QY 2889 -ETPKSYKACSTLPSKNPLDNCPTD--QNKDCKELOTFTFCSKNDYNNLNDNNAYL 2945
 Db 4668 GEVAQKLEAKALD--QAMQALRNSIQOQOTESKFT-----NEDKPOKDAYOA-A 4717
 QY 2946 VLNSSD-----DNKGVLIPRRRLCTRPITAYNRKDKIELK----- 2984
 Db 4718 VONAKDLINQGNPTLDKQVEQLTAQVTTAKDNLHGDQKLARDQQAQVTVNALPRLNH 4777
 QY 2985 --KLLTSAFSQOGL--LQOKYSEEL--CEPAMK-----YSTA----- 3018
 Db 4778 AQOQALDIALNAAPTREVAOHQVOTATELDHAMEYTLNKNKYDOYNTDKAPNYEASTDK 4837
 QY 3019 -----DYSDLIKGTDMMDTSLEKIKKIFETSNATENR-----KTMWENNRQ 3062
 Db 4838 EAYDQALQAAESTIDPTNGSNANKDAVDQVLTTLQEKENELNGNERVAEKKTAQOTIID 4897
 QY 3063 ITHAMLCGYIATSKVTLDEGWCQLPRDEETNOF--LRWLIEWAKQACKEKKHVDSLKT 3120
 Db 4898 LTH--LNAQDIATAKQONIDATKLOPIAELVDAQTLQNGSMQLOQAVNHAVEGTVOY 4955
 QY 3121 KCPRSNEDN-----FASALLRQPGQNDIRKYISINILI-----K 3156
 Db 4956 TQADSDKQNAVKQALIDAEVNLKONANKQOVDQALO--NIINAKQALNGDERVALAKTNGK 5014

QY 3157 NTMENINIKYKOLKQSSGNIDKPKSEENVOSYIKSKDS-----OCALANDINEIYTG 3210
 DB 5015 HOIDQJNTALNNAGQDFKRGIDQSDNLNGIQIYDEAKALNRAMDQISGIEITDNEGRFG 5074
 QY 3211 TKNNEENREKVEUKLYPGIYFEDETHKNHVLGDNIGKEEQIVRKALYFFPHDSFY 3270
 DB 5075 STYIVNADTQ--VKOYV-----DET-----VDKAR 5097
 QY 3271 QADLFTHRAVOYDPKNDILKSSISVIV--SAL--GLIALHFMKKPKSSVDLRILNIPQ 3328
 DB 5098 QALDKST--GQNLTAQYTKLNDAYTAAKKALNGEERLNKRAEALQRLDQTLHNNAG 5154
 QY 3329 GEXGM-----PTLESKNRYI-----PYRSGPKKTYIYMEGDTSG 3364
 DB 5155 ROLAIQOINNAETLNKASRINRATKLDNMGAVOQYIDQHLGVISSFTYINADDLKA 5214
 QY 3365 DEKYMWD-----LSSSDITSESEYEELDINDI-----YRSGPKKTLIEVY 3408
 DB 5215 NYDNATANMAHELDKYQGNIAIAKAEOLKQNTIDQNALNGDQNTANAKKANAFVNSL 5274
 QY 3409 --LEPSKRD-----IPSDPSPNDQPTNRFID--DEMNELKHDPVSQYLPNTEPNKNYK 3459
 DB 5275 NGLNQOQODLAHAINNADIVSDYDVLVNNQIDLNAMETLKH--LVDNELIPNABQYVNTQ 5333
 QY 3460 SADIPIANTEPNTLYSDNPEEKPEIISIHDRDLYTGKEISYINMSTNTNDIPMANRDS 3519
 DB 5334 NAD-----DNAKTNFDD-----AKRLANTLNSDWTN-----VNDI 5364
 QY 3520 YRGIDLINDSLVYN 3534
 DB 5365 NGAIQANDAIHNLN 5379

RESULT 14
 US-10-087-464-10
 : Sequence 10, Application US/10087464
 : Publication No. US20030059436A1
 : GENERAL INFORMATION:
 : APPLICANT: Chishli, Athar
 : APPLICANT: Oh, Steven
 : APPLICANT: Liu, David
 : APPLICANT: Goel, Vikas
 : APPLICANT: Li, Xuerong
 : TITLE OF INVENTION: Band 3 Antigenic Peptides, Malaria Polypeptides and Uses Thereof
 : FILE REFERENCE: S1237/7019
 : CURRENT APPLICATION NUMBER: US/10/087,464
 : CURRENT FILING DATE: 2002-03-01
 : PRIOR APPLICATION NUMBER: US 06/272,930
 : PRIOR FILING DATE: 2001-03-02
 : NUMBER OF SEQ ID NOS: 59
 : SOFTWARE: Patent version 3.0
 : SEQ ID NO 10
 : LENGTH: 1639
 : TYPE: PRT
 : ORGANISM: Plasmodium falciparum
 US-10-087-464-10

Query Match 1.5%; Score 289; DB 9; Length 1639;
 Best local Similarity 19.3%; Pred. No. 1.5e-07;
 Matches 327; Conservative 244; Mismatches 603; Indels 516; Gaps 76;

QY 1650 KGPNNYNNIKELTEDVLPSPRRRLICFHALDGYTDPBEVDENGLRRLMEVATEGYNL 1709
 DB 328 PPPANSQN-----TPNLT-----LDKN-----KKIEHEKEIEMIKTIFKNI 365
 QY 1710 G-----QYKKEKKE-----KITSDAHKIYSTEYPPCSAMKYSEFDLDIIGID 1755
 DB 366 DSLFTDPLELEYLRKKNKIDISAVETKESTEPN--EYENGVTYPLSYNDINNALNELN 1755
 QY 1756 NLBD-----EKQTEENLKIPIKNGTSGKSGSDSTTGPGSTARKPMNEKEEYVWAMIC 1812
 DB 425 SFBDLNPFDYTKPEPSKNITTDN-----EKKKFINETIKETI---KIE 463
 QY 1813 GYKRGDDGNSGNSARSDEDLKCGSVSPDDDPYMGKNRD-----EGTAYQPL--- 1860
 DB 464 KKTIESDKKSYEDRSKSLNDITKEYEKLNEIYDSKRNNDIDLNPEKMMGKRYSYKVEK 523
 QY 1861 ----RWFEMGEDPCFKHELEKLYGACADYTCGDNEDKRRKCTDACTOYKRFISEWKP 1916
 DB 524 LTHHNTFASY--ENSKHNEKLTALKYMEDYSLRNIVSEKE-----LKYKMLISKIEN 576
 QY 1917 QYKQIKKYGKGNKDKIYSEHPVAKDAEDAREYLDKQKICEKSGD--CEYKCMQDVSF 1974
 DB 577 EIEITVENIKKDEQLE--EKKITKD-----ENKPDKEILEVSDIYKVOY 619
 QY 1975 QRLTDGNSQMPASLDEPREVEGKNCQYRGPYRRETPSPRSLSKATASKE-- 2032
 DB 620 QKVLNKKIDELKKTQILKNNVELKHHIHPNSYKQKNOKEPYLLYL-----KKEID 672
 QY 2033 -AKTAPPTKOPKRYENITTEMPAQTTRAAQOTRRKTSTATTESDVGTMKAILSNKP 2091
 DB 673 KLVFMP-----KVESLINEEKKRIKTEGSDSSESTEGEITGOA-----TTKP 717
 QY 2092 DSRG--TEGCPNRTYQYQKMGCIYKSKENNGICMP--PRRKILCINNIOYLYNETEN 2148
 DB 718 GQAGSALBESDYOAOAO-----EQKOAPPPVPPVPEAKAOYPTPPADVNNKTEEN 768
 QY 2149 KRNDIKAEAFIKCAIETQPLMLKYLIIENPAENE--IQNGTIPDEKRIIMYTYGYDKD 2206
 DB 769 VSKLDYLEKLYE--PLMTSYICHKYLIVSHSTNNKILKQYKTKKEESKL--SSCOPLD 824
 QY 2207 MFEGTDISNDKII-----TVTNSVTTILNENKKK-----QDKKD--EELRKT 2249
 DB 825 LLE--NIQNNIPMYSFSDLSNLSOLFMEIYEKEVNCMLYKLDKDKIKNLEAKKY 882
 QY 2250 FWEKKNKFIEMGMIYGLTLYLTDENEKEKIRDMYQVDMKPLPSLEF----- 2298
 DB 883 --STVAKTSSSMQPLS--LTPO--DKPEVASANDTSHSTNLNLSKLFENIISLGKNNK 937
 QY 2299 -----VKRQFLRWFTMAEFCNKRREQLIKLEAGCKEYECNGSNGKTOCEAEAC 2350
 DB 938 IYQELIGQKSEMFYKILKDSDFYNESTNPKVSKAD-----DINSINDESKRKLKEED 993
 QY 2351 V-----TYQ--NFIKWKTEYERQREKFKDKDGKRYKQYPTSTERDIEKATCAHELYM 2402
 DB 994 INKLKTLQSLFDLYNNKTKLER--LFDKKTIVGYK-----MQIKRLTTLKELES 1044
 QY 2403 KL-----KEICGNKDCSCMQKPSOLPKTTQOSQSSDA-----NDMPESLDY 2444
 DB 1045 KLSLNPKNKYLONFVSFFKKEAEIAETENILENTIKILKHRYKGLVYKNGESSPLKT 1104
 QY 2445 VPBEFNKCEPELSKSGSMITHRTKITEPKIPMNCVEKAAVYLSKEANNNDITLKEKFI 2504
 DB 1105 LSEBSIQTEEDNVASLENFVLSKLEGLKONLNEKKKLSYLSGLHH-----L 1153
 QY 2505 IESTKEKESKNSWTNNPCPKPKPYADKTIGRNRCENREENRKYDYWKCYKSKYFY 2564
 DB 1154 IAELEKVIKNNKNTGNSP-----SENNTPVNNALLESYK--KFL 1189
 QY 2565 OEKKRVCVPPRREHMLRNLDEIKRIERLKDSONYLKLVKRT-----ARNEGIDITKFNSE- 2620
 DB 1190 PEGTDVATV-----VSESGSDTLEQSO--PKKPASTHYGAEASNTITTSQNDVDEV 1237

QY 2621 NGCAMPICOTMAYSPADIGDIYRGDMIRIGYLPPEIKLYKVEYIY-----GKMRN 2675
DB 1238 DVIIVPIFESEEDYDILGOVYTGAV--TPSVIINIILSKINEYEYVILKPLAGYRS 1295
QY 2676 KNGRKNKYNDVOYFRSAMMDANRRDIKMTCAIPDADAKLFRGRNDGPERITLIDCKG 2735
DB 1296 LKR--OLENNVMTF-----NANVKDLINS-----RNNKRNKNV--LESDLIP 1335
QY 2736 HKDDPVVDYIPQFRMWTMESEYCKALMELEKFKSCDHCK-----2779
DB 1336 YKDLTSSNYVXPYKFLNK-----EKDKFLSSNYIKSDIDIDINPANDVIGY 1385
QY 2780 ---TSRCKNDYBENK--CEQCTRCQEFKNFYLKAKSLF-----DI-----2816
DB 1386 YKILSKRYSDLSIKRYINDKGENEKLPLFNINIELTYKTVNDKIDLEVIHLAKVLN 1445
QY 2817 ---QSN---KYKEL--YEOPITYKISTYDVHONV-----2843
DB 1446 YTYEKSNEVEKIKELNYLKTIDOKLADFKKNNNFVIGIADLSDYNNNNILTKFLSTGMVF 1505
2844 ---OKLTKFSECSVES--FSEYIHEYSKC---LNYK-----2872
QY 1506 ENLAKTVLSNLDGMLNISOHQCVKQCPONSGCFRHLDERECKCLNLYKQEGDK 1565
DB 1566 CYENPPTNENNGGCDADAKCTEEDSGSNKKITCECTKPDSPYL-----FDG---1614
QY 2923 IQTFECSKNDY 2934
DB 1615 ---IFCSSSNF 1622

RESULT 15
US-09-919-172-29
; Sequence 29, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Paris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919, 172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 29
; LENGTH: 2125
; TYPE: PRF
; ORGANISM: Homo sapiens
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 3774181CD1
US-09-919-172-29

Query March 1.5% Score 287.5; DB 10; Length 2125;
Best Local Similarity 17.8%; Pred. No. 2.6e-07;
Matches 447; Conservative 355; Mismatches 856; Indels 853; Gaps 118;
QY 1240 GSKPRVOIKRKKAEKDECKTVNDI--LKENDGKQVE-----DCHPKKNSNGPYD 1291
DB 19 GKAKITIDLK-----PNSDCPLKTSIPKAIKIDYROIETIYKDDCVLANNSH--RAK 71
QY 1292 WO---CGNINILVEDPRVCM---PPRQKLCVHFLAN--DMEIKKL-----OSOVNLKEA 1337
DB 72 WKVISPTEGNEAMV--PSVCFVPPNKEAV--DLANRIEQOYQVNLVLMHESHNMK-- 124
QY 1338 FIKSAAATFFSWYIKKDEGNELDK-----ELKEGKIPPAFLFSMFTYGGDYDF 1390
DB 125 ---SVSMWYLI-----NEIDIRASNVASIKT--MLPGHQOVLNLSRFEDE 169

QY 1391 L-----FGTDSKSHGEGSKLKEQIDSLFKNGDOKSPNGKTRQEWMTESH-----1436
DB 170 LEDSQESQVSSSDITOLEKENVAVCKQYQELLSKSEREDSESYNNLISVNRIRLRL 229
QY 1437 EIWEMALCALVIRIGAKKDDFTENYGNVYKFSKSTLEEFARPOFLRWLTEMVDDCY 1496
DB 230 ENCEDRLRIQIRTPLEDRDLHESV-----PRI 256
QY 1497 TROKXIKDVQEKCKNDOLKCTEKNCKCEDYKTKMKKKEMIPQDKYKDEBRKRRDR 1556
DB 257 TDEKTLKELEBK--DDGITT---NKEEFSSQAAASS--VPT-----LR 297
QY 1557 QHIGVWVDYGTGNATDYLNRKFTASCG--DKPGSASYVORNIOLEKQAYYADADKHGC 1614
DB 298 SELANVLQ-----MNOYYSMSSTYIDKLTVMNLKNTQAAALVLYIEKLCDEE 348
QY 1615 TFFIENDKTYTSS-----1629
DB 349 EAVIADKNNIENLISTLQWRSEVDEKROVPHALEDELQAKAISDEMFKYKERDLDFO 408
QY 1630 ---KDKCKGLVKEANTGALKWONGKGNPNYNNLKELEDVLPSSRLRICFHALDGYTDP 1687
DB 409 WHEKADQOLV-----RMQNVHVOIDNRDL--EGIGSLIKYRDYHPL---DDW 455
QY 1688 VKDENGKRLRMEVATGEGYNLGOYKREK-----EKIKTSDAHKYSYEVPPCSAMK 1741
DB 456 IQOVETQRIQIENQENSKTLATQNOKMLVSEIKHQSKMEBCQY-----AEQ 507
QY 1742 YSF-----YDLNDIILGIDNDEKQTEENLKKIFNKGTSVKGSDSTGNPGSTARKF 1797
DB 508 YSAIVKDELTMTYRA--MWSQOKSPVKRRRMOSSADLIQEFMDLRT-----555
QY 1798 FWNENKECVWNAICGYRRGDDGNSGNAR--SDEDLKKCGSVSDDYMGKNRDEGT 1855
DB 556 ---RYATALTMYQYIK-----FAGDSLKRLBEETKRC-----KETSHEG 593
QY 1856 AYQFLRMEFAGEDECKHKELEKLVGACNDYTCGDNEDRRKCTDCTAQYKRFISEBK 1915
DB 594 AYSDL-----LOROKATVLE-----NSKLTGKISLEHVALEK 627
QY 1916 POYEQOIKYGENDKIKYSEHPVAKDADEYIDKQKTKCEKKSQDEYCKKDYSTQ 1975
DB 628 KO-----KSRVEELPKVREA--AENELRKOORV-----EDISLQ 661
QY 1976 RLTDGNSQMPASLDDEKVEYEGKNCQVPRGPVRVRETPSPRVSLSKATSKREKAT 2035
DB 662 KIR-----ASEAK-----QYRLE-----TYREKAEREL--690
QY 2036 APPPKOKKVENLTTEHRAQTRTBAA-----OQTRKRSTAVTTSDVGTWVKAI 2086
DB 691 ---ERVQOLITEA--KRAAVEENILNFRNQLBEENTFRRLTEDHL-----732
QY 2087 LSNRPDSRGIEGCPNPTYGQYPRMGCIYGSKENENGICPPRRKTLGINNOY---L 2142
DB 733 ---KRDLSLNDNQOKNL 749
QY 2143 NYETENRNDIKFAITKCAIETQFLMKLYIENPAENELONGTIPDEFKRIPTYG 2202
DB 750 MEELRRRRDNE--ELLKLIKOMEKDLAFQOVAKQKLEKQIELEARRRITTELY 804
QY 2203 DYKMPFEGTDSNDKRIITYY---NSVTIILNENKKKOD-----KKRDE 2244
DB 805 ---TRENALPVCPIQATSCRAVGLQOEHDKQAEIKQOVDDELTAANRAEQ 856
QY 2245 ELRKIFWEKMKKFIWEMITGLTYHLDENKEKIRBNQYNNMTKLTPLSLEFYKRPQF 2304
DB 857 DMRELITTELN-----ALQLEKTSSEKAWL-----LKDKLDEFTNNTLRC 895
QY 2305 LRMETAEFECKNRKKBOLK-----LEAGCKEYEGCNSDQTOEACACACTYQWIF 2358
DB 896 LKLELE-----RKQJAKGYSQQLRELGRQNLQTTGAEMQDASD-----LK 939
QY 2359 KWKTEYRQREKREKDKDGKKYDYDSTERDIEKATCAH-----EYLNMLKELOGNK 2411

Db 940 KIRNTOLESLNHEK-GR-----LQREVDRTTAAHVAEKNIOHLSQHSFDEK 991
 QY 2412 DCS-----CMQRP-----SSQLEPTTQO-----SOSDANDMESLDYVEEFNK-C-ECPELS 2458
 Db 992 ELEHLOJCOGRKSHLQGFESHQLOLQNIKAENNDKIQRLN---ELEKSECHCAEML 1048
 QY 2459 KKSMTHTKITEPKITPMCEVAKAAYLSKAENNDITLKEKFIPIESTKEKSKNSWT 2518
 Db 1049 KOKVEETLRONNETKLMQRIQ-----AESEN---IVLEKQ-----T 1082
 QY 2519 NNNPCDPPKYPADXYIGRRNPPCENREBNPKVDEMKKNSFVQEKKRVCPRRH 2578
 Db 1083 IQORCEALKIQAD-----GFKDQLRSTNEH-----LHQOTEDDF-----ORKI 1122
 QY 2579 MCLRN-----LDEKIERLKDNSNYLLKAVRTANNEGIDILKNFNSGCMANPICD 2630
 Db 1123 KCLBEDLAKSONLYSEK-QKCOQONIIIONTKKEVNLNAEL--NASKKEKRGEOKVQ 1179
 2631 TMYKSPADLAD-IVRGTDMLRIGGYLPPVEIKLYKVEYIYGKWRNKNKGRNK----- 2682
 1180 LQQAQVOELNNRRLKQVDELHLKTEIEQMTIRKKVYLFQESGKFKQSAEPRKMEKIME 1239
 QY 2683-----YNDVOTFRSAMWDANKRDIWKAMTCAPEDAL-----FRKGRMDG 2723
 Db 1240 SKVTENDIGIRLDFVSLQOEN-----SRADENAKLCETNIKELEROLQOYREOMQOG 1293
 QY 2724 FERITLIQDCGHRDDPVD-----DYIPORFMTESEYCCALMELEKFKSCDHC 2778
 Db 1294 QHMEANHYQCOQKREDELIAQREVENLKQKMOQIKEHEQVLVLOCETIK-KSTAKDC 1352
 QY 2779 KTSDRCKNDYDENKCEQCKTRCOEY-----KNFVLKWSLFDIQSNRYKE 2823
 Db 1353-----TFKPDF-----EMTVKECOHSGELSSRNTGHLPTPRSPILRWTQ---EPOPLEE 1399
 QY 2824 LYEQPITKISTYDHOVQNFQKLTFRKSECSVESSEYHETSKCNKFKFENDGSS--- 2880
 Db 1400 KMOHRVVEQIP--KEVO-FOPPGAPLEKESQOCSYSEFSOTSTELQITFEDETPIRSL 1456
 QY 2881-----NIRTYAEET--PKSYKE-ACSTYL-----PSKNPLD 2909
 Db 1457 ELEKIRQALNNSRPVRYQDNCEMELVAVLPLEIAKKNQYDMHTEVTTLKQKNPVP 1516
 QY 2910 NCPIDOKKDCKE---LQTFPCKNDYDNNLNNMAVLYLNSSDN----- 2953
 Db 1517 SA-BEMMEGCRASGKKGDFLKGLEPETFQNFDDHACSVDDEKFEQGLRHVTAR 1575
 2954-----KGVLIIPR-----RRHICTRITAYNRYKGDKE----- 2981
 Db 1576 QLVAKLLDMTTHQLRLGLKTVAEVOKTLNKLTKATSTAGLYLESTKEKISFASAER 1635
 QY 2982 ILKRLTSAFSQO-----LLGOKYKSEEL-----CFEAMK-----YS 3016
 Db 1636 IIDKMAWALAEQAATGFIIDPISQOTYSVEDAVLKGVDPFERIRLLEAKAAYGYS 1695
 QY 3017 YADYSIIKGTDMADTSLSEKIKKIFETSNEATENRKTWMENNRQIMHAMCGYKIATS 3076
 Db 1696 Y-----SKTILSVF-----QAMENRML---DROKGHILEA--QIASG 1728
 QY 3077 KYTLDEGMCOIPKDETNQFRLWLEWAKOACEKKHVSLSLTKCPRSNEDNFEA---S 3133
 Db 1729 GVIDPVGIRVPPPIALQOGL-----LNNALLOFLHEPSSNTRVFP--NPNNKQALYYS 1780
 QY 3134 ELLEPOSCONDIRKYISLNLIKNTMENLNIKYQLKDQSSGNIDNRP-SEENQSYIKS 3192
 Db 1781 ELLEPM--CYFDVESQCLPFPGERNISNLANK-----KTHRISVDTKTSSELTYE--- 1830
 QY 3193 KDSOCALENDINEIYGTAKNENNEKKEVLKKLYPGLIYVEDETHKNHVLG----- 3245
 Db 1831-----AFORNLIERISYLEISGOQOYQWKEAM-----FFESYGHSSHMLTDTKTGLHF 1877
 QY 3246 NIKKEE-EQYVAPKALYFTFTPHVDSFYQAPLESTHRA-----QYDPKNDLKSSISVIV 3299

Db 1878 NINEAIGOTTIDKALV-----KKYQGLITLTELADSLLSRLVPPKD-LHSEVAGYWL 1929
 QY 3300 SALG--LIALHFMKKRKSVDLRLINIPQEGMPTLESKNRYIPRSGPYKGYIY 3357
 Db 1930 TASGERISVILKASBRNLVDRTILRLCLAEQVSTGI-----IDPLGKKRYV 1976
 QY 3358 MEGDTSGDEKYMMDLSSDITSESEYEELDINDIYVPSGPKYKTLIEV 3408
 Db 1977 AEAHLRGLVDEGF-----AQOLRQCELYITIGHPITNKKMSVVEAV 2018

Search completed: April 28, 2003, 10:49:06
 Job time: 156.127 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 8.41989 Seconds
(without alignments)
4367.892 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554
Perfect score: 1543
Sequence: 1 DCHPRKNSNGYPMOCGNIN.....KKEWIPDKRYXDERDKRRF 276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

al number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	154.3	100.0	3542	22	AAB62142
2	634.5	41.1	294	22	AAB62147
3	624	40.4	2703	16	AA770236
4	624	40.4	2710	18	AAW22482
5	624	40.4	2710	21	AA777904
6	624	40.4	3060	18	AAW22475
7	624	40.4	3060	21	AA777905
8	609	39.5	407	22	AAB62148
9	524	34.0	351	22	AAB62149
10	516.5	33.5	308	22	AAB62151

11	493.5	32.0	793	16	AA770234	P. falciparum E31a
12	493.5	32.0	921	18	AAW22480	Plasmodium E31a.
13	493.5	32.0	921	21	AA777902	P. falciparum ebl-1
14	486	31.5	311	22	AAB62150	P. falciparum varC
15	390.5	25.3	2228	20	AAW93944	P. falciparum PFEW
16	382.5	24.8	2182	18	AAW22476	Plasmodium var-1
17	382.5	24.8	2182	21	AA777906	Plasmodium var-1 P
18	359.5	23.3	2913	17	AAW00384	Plasmodium falcipa
19	339.5	22.0	2197	21	AAW18352	Plasmodium falcipa
20	296	19.2	1726	17	AAW00385	Truncated Plasmodi
21	289.5	18.8	1143	23	AA776759	Plasmodium falcipa
22	282	18.3	972	23	AA776761	Plasmodium falcipa
23	274	17.8	350	23	ABW07656	P. falciparum EBA-
24	274	17.8	1435	16	AAW70232	P. falciparum SABB
25	274	17.8	1435	18	AAW22477	Silicic acid bindin
26	274	17.8	1435	21	AA777900	P. falciparum SABB
27	274	17.8	1604	16	AAW70105	TNF-R-EBA 175 fusl
28	274	17.8	1786	14	AAW41043	CD4-EBA175 fusion
29	273.5	17.7	311	21	AA77915	Plasmodium DBL gen
30	261	16.9	1421	23	AAW76764	Plasmodium falcipa
31	260	16.9	616	23	AAW50533	Undifferentiated amln
32	260	16.9	749	16	AAW70233	P. falciparum EBL-
33	260	16.9	749	18	AAW22479	Plasmodium ebl-1
34	260	16.9	749	21	AA777901	P. falciparum ebl-
35	260	16.9	1086	23	AAW76760	Plasmodium falcipa
36	248.5	16.1	1700	21	AAW18144	Plasmodium falcipa
37	231	15.0	302	23	ABW07655	P. falciparum EBA-
38	223	14.5	329	23	ABW07654	P. vivax PYDBP pro
39	223	14.5	1028	14	AAW41044	Plasmodium vivax P
40	223	14.5	1061	16	AAW70231	P. vivax DABP. P1
41	223	14.5	1115	12	AAW13457	Duffy receptor. P
42	223	14.5	1115	18	AAW22478	Duffy antigen bind
43	223	14.5	1115	21	AAW77899	P. vivax DABP bind
44	221.5	14.4	1245	16	AAW70106	TNF-R-P1. vivax Du
45	218.5	14.2	324	21	AAW77911	Plasmodium DBL gen

ALIGNMENTS

RESULT 1	
AAW62142	standard; Protein: 3542 AA.
XX	
XX	AAW62142:
XX	
XX	29-MAY-2001 (first entry)
XX	
XX	P. falciparum FC33, varCSA protein.
XX	
XX	FC33, varCSA protein; chondroitin sulfate A; CSA; var gene; PFEWPI;
XX	erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
XX	malaria; protozoacide.
OS	Plasmodium falciparum.
XX	
XX	W0200116326-A2.
XX	
XX	08-MAR-2001.
XX	
XX	01-SEP-2000; 2000WO-US24195.
XX	
XX	01-SEP-1999; 99US-0152023.
XX	
XX	(USSH) US DEPT HEALTH & HUMAN SERVICES.
XX	
XX	Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
XX	Gysin J, Pouvelle B, Fujii N, Smith J;
XX	
XX	WPI: 2001-235109/24.
XX	
XX	N-PSDB; AAF57301.
XX	
XX	Novel FC33, varCSA protein, useful for modulating parasitized red blood

PT cell binding, sequestration and onset of maternal malaria -
 XX
 PS Claim 12; Page 63-71; 78pp; English.
 CC
 CC The invention relates to a P. falciparum FCRI3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents the P. falciparum FCRI3 varCSA protein.
 CC
 SO Sequence 3542 AA:
 Query Match 100.0%; Score 1543; DB 22; Length 3542;
 Best Local Similarity 100.0%; Pred. No. 8.5e-138;
 Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 1 DCHPKKNSNGYPMOCGNINLVEDPRVCMPPRRQKLCVHFLANDNEIKLQSOVNLKEAF 60
 1279 DCHPKKNSNGYPMOCGNINLVEDPRVCMPPRRQKLCVHFLANDNEIKLQSOVNLKEAF 1338
 QY 61 IKSAAAETFFSWYIYKSKDEGNELDEKELGKIPPAFLSMFYTGDRDFLEGGDISK 120
 DB 1339 IKSAAAETFFSWYIYKSKDEGNELDEKELGKIPPAFLSMFYTGDRDFLEGGDISK 1398
 QY 121 GHGEGSLKQIDSLFRNGDQKSPNGKTRQEWMTSHSHEIWEAMCALVKGAKKDDFTE 180
 DB 1399 GHGEGSLKQIDSLFRNGDQKSPNGKTRQEWMTSHSHEIWEAMCALVKGAKKDDFTE 1458
 QY 181 NYGYNVNVKFSKSTLEEFKRPQFLRLTEWYDDCYTRQKYLKVOEKCKSNDDQKCD 240
 DB 1459 NYGYNVNVKFSKSTLEEFKRPQFLRLTEWYDDCYTRQKYLKVOEKCKSNDDQKCD 1518
 QY 241 TECNKKCEDYVYMKRKKKEMIPQDKYYKDERDKRRF 276
 DB 1519 TECNKKCEDYVYMKRKKKEMIPQDKYYKDERDKRRF 1554
 RESULT 2
 AAB62147
 ID AAB62147 standard; peptide; 294 AA.
 AC AAB62147;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 P. falciparum varCSA polypeptide R29DBL2-gamma.
 FCRI3 varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 KW malaria; protozoasidae; R29DBL2-gamma.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO200116326-A2.
 XX
 PD 08-MAR-2001.
 XX
 PE 01-SEP-2000; 2000WO-US24195.
 XX
 PF 01-SEP-1999; 990S-0152023.
 XX
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujii N, Smith J;
 XX
 DR WPI; 2001-235109/24.
 PT Novel FCRI3 varCSA protein, useful for modulating parasitized red blood
 cell binding, sequestration and onset of maternal malaria -

XX
 PS Disclosure; Page 72; 78pp; English.
 CC
 CC The invention relates to a P. falciparum FCRI3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents a P. falciparum varCSA polypeptide
 CC R29DBL2-gamma.
 CC
 SO Sequence 294 AA:
 Query Match 41.1%; Score 634.5; DB 22; Length 294;
 Best Local Similarity 45.5%; Pred. No. 2.5e-52;
 Matches 131; Conservative 37; Mismatches 91; Indels 29; Gaps 10;
 1 DCHPKKNSNGYPMOCGNINLVEDPRVCMPPRRQKLCVHFLANDNEIKLQSOVNLKEA 59
 DB 12 NCNANKRKN--EWQDKNTEFVGNEGVCMPPRRKSICIHNLTEEQYK---NKYQLEA 66
 QY 60 FKSAAAETFFSWYIYKSKDEGNELDEKELGKIPPAFLSMFYTGDRDFLEGGDIS 119
 DB 67 FKCAKAKETNLMQKYNKNDKNAEEL---LKKGIPEDFMIMETTFEDDFDLENDMG 123
 QY 120 KGHGEGSLKQIDSLFRNGDQKSPNGKTRQEWMTSHSHEIWEAMCALVKGAKKDDFTE 174
 DB 124 K---DVDRYKKNKINVFNNSSKRGFKIDPENWNNENGPQIWNQMLCALIHADKDSIKN 180
 QY 175 KDDFTENYGYNNVVKFSKSTLEEFKRPQFLRLTEWYDDCYTRQKYLKVOE 229
 DB 181 KD---NKYEKVTILAKRDSNGKSTLEEFKRPQFLRLTEWYDDCYTRQKYLKVOE 236
 QY 230 KCKSND--QKCDTECNKCEDYVYMKRKK--EWIPQDKYYKDERDKRRF 274
 DB 237 TCKSIDGQLCKDRGNNKCDYKRYKMKKKEENNLDQKYYKDERDKRRF 284
 RESULT 3
 AAR70236
 ID AAR70236 standard; Protein; 2703 AA.
 AC AAR70236;
 XX
 DT 22-SEP-1995 (first entry)
 XX
 DE P. falciparum Proj3.
 XX
 KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
 KW vaccine.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9507353-A.
 XX
 PD 16-MAR-1995.
 XX
 PE 07-SEP-1994; 94WO-US10230.
 XX
 PF 10-SEP-1993; 93US-0119677.
 XX
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 XX
 DR WPI; 1995-123427/16.
 DR N-PSDB; AAQ83529.
 XX
 PT New erythrocyte binding domain polypeptide(s) - isolated from
 Plasmodium binding proteins, used in diagnosis, treatment and

[illegible]

PT	New malaria vaccines - contains cysteine-rich DBL family protein binding domains homologous domains of the Duffy and sialic acid binding proteins
PS	Disclosure; Page 50-56; 96pp; English.
XX	
CC	This sequence represents Proj3 of Plasmodium. Proj3 belongs to the Duffy binding like (DBL) family of genes which have homology to the Duffy antigen binding protein (DABP) and sialic acid binding protein (SABP) conserved regions (see AA72869 and AA72888 respectively). The var family of genes modulate cytoadherence and antigenic variation of Plasmodium infected erythrocytes. SAMP and the duffy antigen binding protein (DABP) are soluble proteins that appear in the culture supernatant after infected erythrocytes release merozoites. DABP and SABP mediate the binding of merozoites and schizonts to the erythrocyte surface. These proteins are necessary for erythrocyte invasion by the parasite. This sequence can be used in the compositions of the invention. The compositions are for the treatment and prevention of malaria, and comprise either a nucleotide sequence or encoded polypeptide of the var-1, var-2, var-3 or var-7 gene of the DBL gene family, a family of genes having homology with conserved regions of DABP and SABP. The compositions are used for the treatment and prevention of malaria. They are also used in the preparation of vaccines for inducing a protective immune response in a mammal to plasmodium merozoites (especially Plasmodium falciparum or Plasmodium vivax).
CC	
CC	
CC	
CC	
SQ	Sequence 2710 AA;
	Query Match 40.4%; Score 624; DB 18; Length 2710; Best Local Similarity 43.8%; Pred. No. 4.6e-50; Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11
OY	1 DCHPRNSNGYPDWC-GNINLVEDPRVCMPRRROKLCVHFLANDNEIKLOSVMLEKA 59 ::: : :: : ::: : :: :
Db	1367 ECPNRES---YEDMCKNNIDISH-DGACMPRROKLCIYLIAHSQFENITDDNLKA 1422
OY	60 FTKSAAEFTFSWYYKK-K-DGEGLDELKELKGKTPPAFLNSMFTFYDFDLFGTDI 118 ::: : :: : :: :
Db	1423 FKTAFAEFTLSWMQYKSKNDSEAKLDI---GLTSPGSFLTSMWTTFDYRIDICLTNDI 1478
OY	119 SKGHGEGSLRKQIDISLFENGDSKSPNGKTROEMWHESHHEIMCALIVKIGAKKD- 177 : : : : : : : : : : : : : : : :
Db	1479 SKKQNVAAAKOKIGKFESKDGSSSGSLRSQEMWMTNGPELWKGCALTYVTDTDK 1538
OY	178 --FTFNQYGANNKFSDK-STLEEFARPOFLRWLIEWDDCYTRQKYLKDVQEKCS- 233 : : : : : : : : : : : : : :
Db	1539 RKIKNDYSVDKNQONGNPSLEEFAPAKQPLRMWEMEWEECAEGRKENITKDNCNEI 1598
OY	234 NDQLKCDT---ECNRKCDEYVKYMK-KKEWI-----PDQRYYK 268 : : : : : : : : : : : : : : : : : : : : : : : : :
Db	1599 NSTOOCNDAAKHRCNOACRAYOEYVENKKKEFGQTNNFVLKANVGOPDDEYK 1650
RESULT 5	
ID	AA777904
ID	AA777904 standard; Protein; 2710 AA.
XX	AA777904;
XX	
DT	13-JUN-2000 (first entry)
XX	
De	P. falciparum Proj3 binding domain polypeptide.
XX	
KW	DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein; DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; Immunisation; Protozoacide; Proj3.
KW	
XX	Plasmodium falciparum.
OS	
XX	US5993827-A.
XX	
PN	30-NOV-1999.
PD	
XX	

PF 07-JUN-1995; 95US-0487826.
 PR 10-SEP-1993; 93US-0119677.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
 PI WPI; 2000-194198/17.
 DR N-PSDB; AAZ98286.
 XX
 XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX
 XX Disclosure; Columns 79-92; 93pp; English.
 PS
 XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria. The present sequence represents the Proj3 binding domain
 CC polypeptide.
 XX
 XX Sequence 2710 AA;
 SQ
 Query Match 40.4%; Score 624; DB 21; Length 2710;
 Best Local Similarity 43.8%; Pred. No. 4.6e-50;
 Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;
 QY 1 DCHPKNSNGYPMQC-GNINIVEDPRVCMPPRRORICVHFLANDNEIKKLSOVNLKEA 59
 DB 1367 ECHPKES---YPMDCCKNNIDISHD-GACMPPRORICLYIAHESQTEIKITDNLKDA 1422
 QY 60 FIKSAAEFFFSWYKSK-DGEGNLDKELKGGKIPPAFLRSMYTFEDYDFLGTDI 118
 DB 1423 FIKTAAAEFTLSMOKYKSKNDSEAKILDR---GLIPSOFLRSMYTFEDYDFLGTDI 1478
 QY 119 SKHGEGSKLKEQIDSLFKNGDQKSPNGKTRQEMWTEHSEIWEALCALVIGAKKDD- 177
 DB 1479 SKKQNDVAKAKDKIGKFEKDGSKSPSGLSRQEMWTEHSEIWEALCALVIGAKKDD- 1538
 178 --FTENGYNNVYKFSK-STILEEPAKRPQFLRWLTETWYDDCYTRQKYLKDQEKCKS- 233
 1539 RKIKNDYSIDKVNQSONGNSPSEEFPAKRPQFLRWMIEMGEFCAEROKENIKDACNEI 1598
 QY 234 NDQKCDT---ECNKKCEDYVYKMK-KKKEWT-----PODKYYK 268
 DB 1599 NSTQOCNDAKKHCNOACRAIYQETVENKKKEFSGQTNNEVLKANVOPDPEYK 1650
 RESULT 6
 AAM22475
 ID AAM22475 standard; Protein: 3060 AA.
 AC AAM22475;
 XX
 XX 12-SEP-1997 (first entry)
 DT
 XX Plasmodium var-7.
 DE
 XX Plasmodium var-7.
 XX DBL gene family; SABP: stialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP, merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 XX Plasmodium vivax.
 OS

OS Plasmodium falciparum.
 XX
 XX WO9640766-A2.
 PN
 XX 19-DEC-1996.
 PD
 XX
 XX 07-JUN-1996; 96WO-US09508.
 PF
 XX 07-JUN-1995; 95US-0487826.
 PR
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI; 1997-052231/05.
 DR N-PSDB; AAT72882.
 XX
 XX New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stialic acid
 PT binding proteins
 XX
 XX Claim 8; Page 61-67; 96pp; English.
 PS
 XX This sequence represents var-7 of Plasmodium. Var-7 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stialic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family. A
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC They are also used in the preparation of vaccines and prevention of malaria.
 CC (especially Plasmodium falciparum or Plasmodium vivax).
 XX
 XX Sequence 3060 AA;
 SQ
 Query Match 40.4%; Score 624; DB 18; Length 3060;
 Best Local Similarity 43.8%; Pred. No. 5.4e-50;
 Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;
 QY 1 DCHPKNSNGYPMQC-GNINIVEDPRVCMPPRRORICVHFLANDNEIKKLSOVNLKEA 59
 DB 1365 ECHPKES---YPMDCCKNNIDISHD-GACMPPRORICLYIAHESQTEIKITDNLKDA 1420
 QY 60 FIKSAAEFFFSWYKSK-DGEGNLDKELKGGKIPPAFLRSMYTFEDYDFLGTDI 118
 DB 1421 FIKTAAAEFTLSMOKYKSKNDSEAKILDR---GLIPSOFLRSMYTFEDYDFLGTDI 1476
 QY 119 SKHGEGSKLKEQIDSLFKNGDQKSPNGKTRQEMWTEHSEIWEALCALVIGAKKDD- 177
 DB 1477 SKKQNDVAKAKDKIGKFEKDGSKSPSGLSRQEMWTEHSEIWEALCALVIGAKKDD- 1536
 178 --FTENGYNNVYKFSK-STILEEPAKRPQFLRWLTETWYDDCYTRQKYLKDQEKCKS- 233
 1537 RKIKNDYSIDKVNQSONGNSPSEEFPAKRPQFLRWMIEMGEFCAEROKENIKDACNEI 1596
 QY 234 NDQKCDT---ECNKKCEDYVYKMK-KKKEWT-----PODKYYK 268
 DB 1597 NSTQOCNDAKKHCNOACRAIYQETVENKKKEFSGQTNNEVLKANVOPDPEYK 1648
 RESULT 7
 AAT77905

ID AAY77905 standard; protein; 3060 AA.
XX
AC AAY77905;
XX
DT 13-JUN-2000 (first entry)
XX
DE Plasmodium var-7 polypeptide.
XX
KM DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
XX protozoacide; var-7.
XX
OS Plasmodium sp.
XX
PN US5993827-A.
XX
PD 30-NOV-1999.
XX
PF 07-JUN-1995; 95US-0487826.
XX
XX 10-SEP-1993; 93US-0119677.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Slim KL, Chittis C, Peterson DS, Su X, Wellens TE, Miller LH;
XX
DR WPI; 2000-194198/17.
XX
DR N-PSDB; AA298287.
XX
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria -
XX
XX Disclosure; Columns 109-124; 93pp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially
XX identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
XX Binding Protein (SABP), which are soluble proteins that appear in the
XX culture supernatant after erythrocytes infected with malaria release
XX merozoites. Immunochemical studies indicate that DABP and SABP are the
XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
XX and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
XX used to vaccinate against malaria, especially caused by P. falciparum.
XX Immunization with the polypeptide provides effective protection against
XX malaria. The present sequence represents the var-7 polypeptide.
XX
SO Sequence 3060 AA;
XX
Query Match 40.4%; Score 624; DB 21; Length 3060;
Best Local Similarity 43.8%; Pred. No. 5.4e-50;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;
XX
OY 1 DCHPRKNSNGYPMOC-GNINLVEDPRVCMPPRRKLCYHFLANDNEIKKLSQVNLKEA 59
DB 1365 ECHPKES---YPMDCNNIDISHD-GACMPPRKLCYHFLANDNEIKKLSQVNLKEA 1420
XX
OY 60 FIKSAAEETFFSWYYKSK-DSEGNELDKELKGIKIPPAFLRSMFYFGDYRDLFGTDI 118
DB 1421 FIKTAAAEFTLSWYKSKNDSEAKILDR---GLIPSGFLRSMWTFGDIYRDLFGTDI 1476
XX
OY 119 SKHGGSKLKEQIDSLFKNGDQSPNGKTRQWTEHSHETIEMALCALVTKGAKKD- 177
DB 1477 SKRQNVAAKAKDKIGFFSKDGSKSPSGLSRQWMTKNGEIKWGMICALTKVVTDDNK 1536
XX
OY 178 --FTENGYNNVSKF--STTLEEFARPOFLMLTEWYDGYCYTRQYKLDVOEKCS- 233
DB 1537 RRTKNDYSIDKVNQSONGNSTLEEFARPOFLMLTEWYDGYCYTRQYKLDVOEKCS- 233
XX
OY 234 NDLAKCDT---ECNKCEYDVYKMK-KKKEMI-----PDKYK 268
DB 1597 NSTQCCNDAKHRCNQACRAVOYEYENKKEFSGOTNNFVLKAVNQPDPEYK 1648

RESULT 8
AAB62148
ID AAB62148 standard; peptide; 407 AA.
XX
AC AAB62148;
XX
DT 29-MAY-2001 (first entry)
XX
DE P. falciparum varCSA polypeptide A4 DBL4-gamma.
XX
KM FCRI3 varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
KW erythrocyte membrane protein 1; parasitized red blood cell; PIRC;
XX malaria; protozoacide; A4 DBL4-gamma.
XX
OS Plasmodium falciparum.
XX
PN WO200116326-A2.
XX
PD 08-MAR-2001.
XX
PF 01-SEP-2000; 2000WO-US24195.
XX
PR 01-SEP-1999; 99US-0152023.
XX
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheideid C;
XX Gysin J, Pouvelle B, Fujii N, Smith J;
XX
DR WPI; 2001-235109/24.
XX
XX Novel FCRI3 varCSA protein, useful for modulating parasitized red blood
XX cell binding, sequestration and onset of maternal malaria -
XX
XX Disclosure; Page 72-73; 78pp; English.
XX
XX The invention relates to a P. falciparum FCRI3 varCSA protein, that is
XX capable of binding to chondroitin sulfate A (CSA). The var gene and the
XX corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1)
XX modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
XX protein and the encoding gene are useful for treating and preventing
XX maternal malaria in a patient identified at a risk for contracting
XX maternal malaria or in a patient afflicted with maternal malaria. The
XX present sequence represents a P. falciparum varCSA polypeptide
XX A4 DBL4-gamma.
XX
SO Sequence 407 AA;
XX
Query Match 39.5%; Score 609; DB 22; Length 407;
Best Local Similarity 42.3%; Pred. No. 1.1e-49;
Matches 127; Conservative 50; Mismatches 89; Indels 34; Gaps 10;
XX
OY 1 DCHPRK--KNSNGYPMOCGNINLVEDPRVCMPPRRKLCYHFLANDNEIKKLSQVNLKE 58
DB 11 ECHPKIYPTNDPGWNTDQKVNIREGSCMPRRKLCYHFLANDNEIKKLSQVNLKE 67
XX
OY 59 AFIKSAAEETFFSWYYK-----SKDEG-----NELDKELKGIKIPPAFLRSMFYTE 106
DB 68 AFIECAAEETFFSWYYK-----SKDEG-----NELDKELKGIKIPPAFLRSMFYTE 106
XX
OY 107 GDYRDLFGTDISKGHGGSKLKEQIDSLFKNGDQSPNGKTRQWTEHSHETIEMALC 166
DB 128 GDYRDLFGTDISKGHGGSKLKEQIDSLFKNGDQSPNGKTRQWTEHSHETIEMALC 166
XX
OY 167 AL---VRIKAKDPTFNQYNNVSKFSDKSTTLEEFARPOFLMLTEWYDGYCYTRQYK 223
DB 188 GLSHHILN--NKQDLRKNLNDN--KYTKISSKLEDFASRQFLMLTEWYDGYCYTRQYK 245
XX
OY 224 LKDVQEKSKNDLQCDTE-----CNKCEYDVYKMKKKKEMIPOD--KYKDEBDK 274
DB 246 INDLKTC---NNEEGSGQNGKREKCKNACENAVKSLKMKQDYEQDTAKFKDKDKK 302

RESULT 11
 AAR70234 standard; Protein; 793 AA.
 ID AAR70234
 XX AAR70234;
 AC
 XX 22-SEP-1995 (first entry)
 DT
 XX P. falciparum E31a.
 DE
 XX Erythrocyte binding ligand; E31a; binding domain; malaria; therapy;
 KM vaccine.
 KM Plasmodium falciparum.
 OS
 XX WO9507353-A.
 PN
 XX 16-MAR-1995.
 PD

07-SEP-1994; 94WO-US10230.
 PR 10-SEP-1993; 93US-0119677.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 PI WPI: 1995-123427/16.
 DR N-PSDB: AAO83527.
 XX New erythrocyte binding domain polypeptide(s) - isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and
 PR prevention of malaria
 XX
 PS Disclosure: Page 51-52; 81pp; English.
 XX
 CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AAO83526), E31a (AAO83527), EBL-e2 (AAO83528) and Pro3
 CC (AAO83529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 CC
 XX

SO Sequence 793 AA:

Query Match 32.0%; Score 493.5; DB 16; Length 793;
 Best Local Similarity 40.0%; Pred. No. 2,8e-38;
 Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPRKNSNGYPPMOCGNIL-VEDPRVCMPPRROKLCVHFLANDNEIKLOSQV-LKE 58
 DB 414 EYRKRT--YSEWTCDESKIKMGHGACIPPRROKLCVHFL--EKIMTNTELK 464
 QY 59 APTKSAATETFSWY-VYKSKDEGNELDKELKGIPTAFILSMFTYTGVDYDFLFGTD 117
 DB 465 AFITKAAAEFTFLMOMYKKDKGNMADDEKLKGLIIPDEFKQMTYTTADYDIDLGTD 524
 QY 118 I-----SKGHGSKLQKQIDSLFNKGQKSPNG-KTROEMWTEHSHEIWEAMLCAL-- 168
 DB 525 ISSKMDTSKGVG---KYKCNIDVF---YKISNSIRYKRSWMTNGPVIWEGMLCALSY 577
 QY 169 -----VKIGAKKDPTEVNGYNNVKE-SDKSTLEEFAPKPOFLMFLWEDYDCTTR 221
 DB 578 DTSINAVNETAKKLTGNNNEKAVTFGSDSTLTLSKFSERPOFLMFLWEGENFEKOR 637
 QY 222 KYLDVQERCKNSDQKCTECNKK---CEDYVK-----YMKKKKEMIQDR- 265
 DB 638 KEKVVLLACKDCD-VDGQKCKGKCVACKDCKQYHSMIGIWIWIDYKKQKGYTEVKKI 696

QY 266 -YKDERDKK 274
 DB 697 PLYKEDKDVK 706

RESULT 12
 AAM22480 standard; Protein; 921 AA.
 ID AAM22480
 XX AAM22480;
 AC
 XX 07-OCT-1997 (first entry)
 DT
 XX Plasmodium E31a.
 DE
 XX DBL gene family; SABB; stailic acid binding protein; vaccine; therapy;
 KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KM Plasmodium.
 KM
 XX Plasmodium falciparum.
 OS
 XX WO9640766-A2.
 PN
 XX 19-DEC-1996.
 PD
 XX 07-JUN-1996; 96WO-US09508.
 PF
 XX 07-JUN-1995; 95US-0487826.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Chitnis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 PI WPI: 1997-052231/05.
 DR N-PSDB: AAT72893.
 XX

New malaria vaccines - contains cysteine-rich DBL family protein
 binding domains homologous domains of the Duffy and stailic acid
 binding proteins
 PS Disclosure: Page 43-45; 96pp; English.

This sequence represents E31a of Plasmodium. E31a belongs to
 the Duffy binding like (DBL) family of genes which have homology to the
 Duffy antigen binding protein (DABP) and stailic acid binding protein
 (SABB) conserved regions (see AAT72889 and AAT72888 respectively). The
 var family of genes modulate cytoadherence and antigenic variation of
 Plasmodium infected erythrocytes. SABB and the Duffy antigen binding
 protein (DABP) are soluble proteins that appear in the culture
 supernatant after infected erythrocytes release merozoites. DABP and
 SABB mediate the binding of merozoites and schizonts to the erythrocyte
 surface. These proteins are necessary for erythrocyte invasion by the
 parasite. This sequence can be used in the compositions of the
 invention. The compositions are for the treatment and prevention of
 malaria, and comprise either a nucleotide sequence or encoded polypeptide
 of the var-1, var-2, var-3 or var-7 genes of the DBL gene family. a
 family of genes having homology with conserved regions of DABP and SABB.
 The compositions are used for the treatment and prevention of malaria.
 They are also used in the preparation of vaccines for inducing a
 protective immune response in a mammal to Plasmodium merozoites
 (especially Plasmodium falciparum or Plasmodium vivax).

SO Sequence 921 AA:

Query Match 32.0%; Score 493.5; DB 18; Length 921;
 Best Local Similarity 40.0%; Pred. No. 3.4e-38;
 Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPRKNSNGYPPMOCGNIL-VEDPRVCMPPRROKLCVHFLANDNEIKLOSQV-LKE 58
 DB 414 EYRKRT--YSEWTCDESKIKMGHGACIPPRROKLCVHFL--EKIMTNTELK 464

QY 59 AFKSAAEFFSWY-YRKSNDGEGNELDKELKEKIPAPFLRSMFYFGYRDLFGTD 117
 DB 465 AFKCAAAETFLMKNYKDKNGNAEDDEKLGKGIIPEDFRKQRYFYFADYRICIGTD 524
 QY 118 I-----SKGHEGSKLEQIDSLFKNGDQSPNG-KTROEMWTEHSHIEWAMICAL-- 168
 DB 525 ISSKDTSGVG--KVCNIDDFV---YKISNSIRYKSMWETNGFVWEGMICALSY 577
 QY 169 -----VKIGAKKDDFTENYGYNNYK-FSDKSTLEEFARKPOFLRWLTWYDYCYTRQ 221
 DB 578 DTSLNANVPETHKKLTBEGNNNEFYIGSDSTLSKFSERPOFLRWLTWEGENFCKEOK 637
 QY 222 KYLDVQEKCKSNDQKCDTECNK---CEDYK-----YMKKKKEWIPQDK- 265
 DB 638 KEYVTLAKCKDCD-VDGDKNGKCVACKQCKQYHSMIGIWDNKKQKGRTEYVKI 696
 QY 266 -YYKDERDK 274
 DB 697 PLYKEDKDVK 706

RESULT 13
 AAY77902
 ID AAY77902 standard; Protein; 921 AA.
 AC AAY77902;
 DT 13-JUN-2000 (first entry)
 DE P. falciparum ebl-1 related polypeptide.
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KM DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
 XX Protozoacide.
 OS Plasmodium falciparum.
 XX US5993827-A.
 PN 30-NOV-1999.
 PD 07-JUN-1995; 95US-0487826.
 PF 10-SEP-1993; 93US-0119677.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Slim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
 DR MPI; 2000-194198/17.
 N-PSDB; AA298284.

Isolated protein binding domains from Plasmodium vivax and Plasmodium falciparum erythrocyte binding proteins useful for vaccinating against malaria -
 Disclosure: Columns 61-66; 93pp; English.

The invention relates to ebl-1 polypeptides that are encoded by the DBL (Duffy-binding like) gene family. The ebl-1 proteins are substantially identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid Binding Protein (SAMP), which are soluble proteins that appear in the culture supernatant after erythrocytes infected with malaria release merozoites. Immunochemical studies indicate that DABP and SAMP are the respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be used to vaccinate against malaria, especially caused by P. falciparum. Immunization with the polypeptide provides effective protection against malaria.

Sequence 921 AA;

Query Match 32.0%; Score 493.5; DB 21; Length 921;
 Best Local Similarity 40.0%; Pred. No. 3.4e-38;
 Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPRKNSNGYIPDMOGNINL-VEDPVCMPPRRKCLCVHFLANDNEIKKLQSYN-LKE 58
 DB 414 ECTRKET---YSEWTCDESKIRKMGHACIPPRKLCIHYL-----EKIMTNTNELKY 464

QY 59 AFKSAAEFFSWY-YRKSNDGEGNELDKELKEKIPAPFLRSMFYFGYRDLFGTD 117
 DB 465 AFKCAAAETFLMKNYKDKNGNAEDDEKLGKGIIPEDFRKQRYFYFADYRICIGTD 524
 QY 118 I-----SKGHEGSKLEQIDSLFKNGDQSPNG-KTROEMWTEHSHIEWAMICAL-- 168
 DB 525 ISSKDTSGVG--KVCNIDDFV---YKISNSIRYKSMWETNGFVWEGMICALSY 577
 QY 169 -----VKIGAKKDDFTENYGYNNYK-FSDKSTLEEFARKPOFLRWLTWYDYCYTRQ 221
 DB 578 DTSLNANVPETHKKLTBEGNNNEFYIGSDSTLSKFSERPOFLRWLTWEGENFCKEOK 637
 QY 222 KYLDVQEKCKSNDQKCDTECNK---CEDYK-----YMKKKKEWIPQDK- 265
 DB 638 KEYVTLAKCKDCD-VDGDKNGKCVACKQCKQYHSMIGIWDNKKQKGRTEYVKI 696
 QY 266 -YYKDERDK 274
 DB 697 PLYKEDKDVK 706

RESULT 14
 AAB62150
 ID AAB62150 standard; peptide; 311 AA.
 AC AAB62150;
 DT 29-MAY-2001 (first entry)
 DE P. falciparum varCSA polypeptide FC3 var3DBL-gamma.
 KW FC3 varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; FC3 var3DBL-gamma.
 OS Plasmodium falciparum.
 XX WO200116326-A2.
 PN 08-MAR-2001.
 PD 01-SEP-2000; 2000WO-US24195.
 PF 01-SEP-1999; 99US-0152023.
 PR (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujii N, Smith J;
 DR MPI; 2001-235109/24.

Novel FC3 varCSA protein, useful for modulating parasitized red blood cell binding, sequestration and onset of maternal malaria -
 Disclosure: Page 74-75; 78pp; English.

The invention relates to a P. falciparum FC3 varCSA protein, that is capable of binding to chondroitin sulfate A (CSA). The var gene and the corresponding P. falciparum erythrocyte membrane protein 1 (PFEMP1) modulate adhesion of parasitized red blood cell (PRBC) to CSA. The protein and the encoding gene are useful for treating and preventing maternal malaria in a patient afflicted with a risk for contracting maternal malaria or in a patient afflicted with maternal malaria. The present sequence represents a P. falciparum varCSA polypeptide

```

CC blood cells, this prevents occlusion of capillaries as is seen in
CC cerebral malaria caused by Plasmodium falciparum. The products of the
CC invention can be used to treat malaria or to vaccinate against it, or
CC used to design a model to identify compounds that bind to PfEMP1. The
CC carbohydrates, polypeptides and antibodies of the invention can be used
CC as a medication for dissolving the rosettes formed by erythrocytes
CC infected by a malaria parasite. The products can also be used for
CC detection, diagnosis and drug screening.
XX
XX Sequence 2228 AA;
XX
XX Query Match 25.3%; Score 390.5; DB 20; Length 2228;
XX Best Local Similarity 28.3%; Pred. No. 7,7e-28;
XX Matches 113; Conservative 48; Mismatches 103; Indels 135; Gaps 18;
XX
QY 8 SNGYEDWOC---GNINIVEDPR--VCMPPRRQKLCVHFL-----ANDNEIKKL 50
DB 911 NNSRLGHCVCYPPSGEPTTSSDKNGAICVPPRRRLYIKKIYDMAKTESPQASGSEAST 970
QY 51 -----QSQVNLKAEAFKSAAEFFEFWYTK-----SKDGBGNEL----- 85
DB 971 SGSTPPDSKELALKAFVESAAIEFFLWHRKKEKKAVALAOGAGHGLPVEEGSPEDP 1030
QY 86 DKELEKGKIPAPFLRSMFYTFGDYRDFLFG-----TDISK-----GHGE 124
DB 1031 EDKLEGGKIPDGFLLKQMFYTLGDYRDLFLSGSNDTTSVSKDTPSSSDNLKNIYLLASG 1090
QY 125 GSKLKEQIDSL--FKN-----GDQKSPNGKTR-QEMTTEHSHEIWEAMICALV--KIG- 172
DB 1091 TGEDEKKNKKYKEIKNFKRCSTERSAPNLVSHPTWMENNGKYLIMHGVCALTSKDKIAK 1150
QY 173 --AKKDFETN-----YGVNNKFSDKS-----TTLE 197
DB 1151 GVEKKPKKIEPNENIMDEANKKPKRPQYQYTFVKLDENSGTSPRTTQTOASSDNTPTTIL 1210
QY 198 EFARKPQFLRWLTETRYDDYCTRQKYLADVDQEKK-----SNDQKQDT----- 241
DB 1211 HFVKRPPTYFRWFEEMGESFCRRERKRLKQIKVDCXKVENGDVGRCSGDGEACDSISTHDS 1270
QY 242 -----ECNKKCEDYVKYMKKK--EWLPQDKYYKDER 271
DB 1271 TVPSFNCPCGCGKHCSSYRKWIERKKIEHHKQSNAYGQOK 1309

```


GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 5.10497 Seconds

(without alignments)
5197.501 Million cell updates/sec

Title: US-10-087-013-2-COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPRKNSNGYPMQGNIN.....KKWIPDKYKDERDKRRE 276Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000Post-processing: Minimum Match 08
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	634.5	41.1	2706	2 T28155	variant-specific s
2	624	40.4	3078	2 T28432	variant-specific s
3	609	39.5	3026	2 T28431	variant surface pr
4	486	31.5	3006	2 T28625	variant-specific s
5	475	30.8	1711	2 C71625	variant-specific s
6	410.5	26.6	2042	2 T18399	variant-specific s
7	402.5	26.1	2135	2 T14602	variant-specific s
8	396	25.7	2212	2 T28157	variant-specific s
9	390.5	25.3	2228	2 T14029	erythrocyte membra
10	385.5	25.0	2664	2 T28626	variant-specific s
11	382.5	24.8	2482	2 T28634	variant-specific s
12	371.5	24.1	2647	2 T28161	variant-specific s
13	369.5	23.9	2924	2 T18378	variant-specific s
14	339.5	22.0	2197	2 B71600	variant-specific s
15	299.5	19.4	1729	2 T18396	erythrocyte membra
16	274	17.8	1435	2 A37793	erythrocyte bindin
17	236	15.3	1153	2 T28652	erythrocyte bindin
18	223.5	14.5	1045	2 T18373	erythrocyte bindin
19	223	14.5	1070	2 T30848	Duffy receptor - p
20	172.5	11.2	778	2 A35970	erythrocyte-bindin
21	106	6.9	3119	2 T18414	protein g377 - mal
22	105.5	6.8	346	2 D64484	hypothetical prote
23	104.5	6.8	378	2 T18486	hypothetical prote
24	104.5	6.8	781	2 F90547	hypothetical prote
25	103.5	6.7	677	2 B29959	DNA-directed RNA p
26	101	6.5	522	2 C96608	hypothetical prote
27	100.5	6.5	980	2 E71606	hypothetical prote
28	100	6.5	227	2 D90512	conserved hypothet
29	99.5	6.4	759	2 D71853	hypothetical prote

30	99	6.4	1650	2 T18444	hypothetical prote
31	99	6.4	3660	1 S02041	dystrophin, muscle
32	98.5	6.4	1345	2 A64430	DNA-directed RNA p
33	98.5	6.4	1467	2 T23950	hypothetical prote
34	98	6.4	2010	2 B71616	phosphatase (acid
35	97	6.3	963	2 C90535	conserved hypothet
36	96.5	6.3	938	2 G70472	hypothetical prote
37	95.5	6.2	37	2 A54514	glutamic acid-rich
38	95	6.2	1391	2 T20642	hypothetical prote
39	95	6.2	1397	2 E87998	protein F09C3.111
40	95	6.2	4427	2 P0637	polyketide synthas
41	94.5	6.1	446	2 A48407	neutral class-II be
42	94.5	6.1	1178	2 S78475	mannosylphosphoryl
43	94	6.1	1081	1 A42399	isooleucine-tRNA 11
44	93.5	6.1	612	2 C72293	hypothetical prote
45	93	6.0	428	2 A45568	highly immunoreact

ALIGNMENTS

```
RESULT 1
T28155
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragne
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
C:Accession: T28155
R:Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.
Nature 388, 292-295, 1997
A:Title: Plasmodium falciparum reosetting is mediated by PfEMP1 and requires complemen
A:Reference number: Z20477; PMID:97373957; PMID:9230440
A:Accession: T28155
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2706 <ROW>
A:Cross-references: EMBL:Y13402; PIR:CAA73831.1
A:Experimental source: strain IT 4/25/5
C:Genetics:
A:Introns: 2493/3
A>Note: R29R+var1

Query Match      41.1%; Score 634.5; DB 2; Length 2706;
Best Local Similarity 45.5%; Pred. No. 1.2e-39;
Matches 131; Conservative 37; Mismatches 91; Indels 29; Gaps 10;

QY 1 DCHPRKNSNGYPMQGNINLVE-DPRVCMPPRRQKCYHFLANDNIRKLOQVYNKEA 59
DB 787 NCNAKNRKN-EMQCDKNTFVGNEGVCMPPRRKSICINLTLEQTK--NKYQLREA 841
QY 60 FIKSAAEFFPSWYVYKSGDGEKELKEGKIPAPFMSMPYFGDRDFLFGDIS 119
DB 842 FIKCAKETMLMDKYKNDNEAEEL--LAKGKIPDFMRIFETGDRDRCLENDMG 898
QY 120 KGHGEGSKLEQIDSLFKNDQKSPNGKTRQEWTEHSHIWEMLCALVYKIGAK---- 174
DB 899 K--DVDKVKKNINKVFNNSSKRGFKKIDPENWNNENGPOIMNGMICALIHADTKSDIKN 955
QY 175 KDOFTENYGYNNVNFSDK-----STLEEFAPKPPFLRMLEWYDDVCYRQYKIQVE 229
DB 956 KD---NYKTEKVTILAKRGSNGMTSEFAKKYFLRMVEVDDDKCKRQYLTVEVAS 1011
QY 230 KCKSND--QLKCDTECNKCCEDYVYKMKKKR-EMIPDKYKDERDK 274
DB 1012 TCKSIDGQLKCDRCGNCKNCKDEYKRYKRRKKEEMNLQDKYKDKREMK 1059

RESULT 2
T28432
variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
```

C:Accession: T28432
 R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, D.S.
 Cell 82, 89-100, 1995
 A:Title: The large diverse gene family var encodes proteins involved in cytoadherence an
 A:Reference number: 220487; MUID:95330813; PMID:7606788
 C:Accession: T28432
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3078 <SUX>
 A:Cross-references: EMBL:LA0608; NID:9886374; PID:9886375; PIDN:AA75396.1
 C:Genetics:
 A:Gene: var-1
 A:Introns: 2611/3

Query Match 40.4%; Score 624; DB 2; Length 3078;
 Best Local Similarity 43.8%; Pred. No. 8.5e-39;
 Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPKNSNGYPMOC-GNINLVDPVCMPPRRKQICVHFLANDNEIKKLOSOVNIKEA 59
 1365 ECPNPKES---YPMDCNNIDISHD-GACMPPRKQICLYIAHESQTENIKIKTDMLKDA 1420
 60 FIKSAAEFTFSWYKSK-DGEGNELDKELKGIKIPPAFLRSMFTFGDYRDLFGTDI 118
 DB 1421 FIKTAAEFTLSWQYKSKNDSEAKILDR---GLIPSOFLRSMATTFEDYRDLCLNTDI 1476
 QY 119 SKHGSGSKLEQIDSLFKNGDQSPNGKTRQWTEHSHHEWAMICALVIGAKKDD- 177
 DB 1477 SKQNDVAKAKDKIGKFFSKDQSGSPGLSRQWMTKNGPELWIKMICALTKYVDTDNK 1536
 QY 178 -FTENYGYNNVFKSDK-STTLEEFARPOFLRWLEWYDDCYTRQKLYKQVECKRS- 233
 DB 1537 RKIKNDYSYDKVQNGNPSLEFAKQFLRMLEKGEBCAERKQKENTIKDKCNEI 1596
 QY 234 NQDLKCDT---BNKCEIYVYKMK-KKKEWT-----PODKYK 268
 DB 1597 NSTQQNDAKHRCNACRAIOEYVENKKKEFGQTNNEFLYLANVQPODEYK 1648

RESULT 3

T28431

variant surface protein 1 homolog AAVAR - malaria parasite (Plasmodium falciparum) (frag
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28431
 R:Smith, J.D.; Kyes, S.; Craig, A.G.; Fagan, T.; Hudson-Taylor, D.; Miller, L.H.; Baruch
 Mol. Biochem. Parasitol. 97, 133-148, 1998
 A:Title: Analysis of adhesive domains from the AAVAR Plasmodium falciparum erythrocyte m
 Reference number: 220486; MUID:99094502; PMID:9879893
 C:Accession: T28431
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3026 <SMI>
 A:Cross-references: EMBL:LA2244; NID:93540144; PID:93540145; PIDN:AA03351.1
 C:Genetics:
 A:Gene: var
 A:Introns: 2906/3

Query Match 39.5%; Score 609; DB 2; Length 3026;
 Best Local Similarity 42.3%; Pred. No. 1.1e-37;
 Matches 127; Conservative 50; Mismatches 89; Indels 34; Gaps 10;

QY 1 DCHRK--KNSNGYPMOCGNINLVDPVCMPPRRKQICVHFLANDNEIKKLOSOVNIKE 58
 DB 2005 ECPNPKES---YPMDCNNIDISHD-GACMPPRKQICLYIAHESQTENIKIKTDMLKDA 2061
 QY 59 AFISAAAEFTFSWYKSK-DGEGNELDKELKGIKIPPAFLRSMFTFGDYRDLFGTDI 106
 DB 2062 AFISAAAEFTFSWYKSK-DGEGNELDKELKGIKIPPAFLRSMFTFGDYRDLFGTDI 118
 QY 107 GDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFG 166
 DB 2122 GDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFGDYRDLFG 2181

QY 167 AL--VKIGAKDDFTENGYNNVFKSDK-STTLEEFARPOFLRWLEWYDDCYTRQK 223
 DB 2182 GLSHHNTKNG-NKQOLKRLNLTNDN-KYTKISSKLEDFASNPQLRWLEWYDDCYTRQK 2239
 QY 224 LKQVQCKSKNDQKCDTE-----CNKCEDYKLYKMKKKEMIPQD--KYKDERDK 274
 DB 2240 INQKTCG---NEYECGSENGKCKKCNACAYRSMWLDKWDQYEOQTAKFDKDKDK 2296

RESULT 4

T28625

variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T28625
 R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson,
 Cell 82, 89-100, 1995
 A:Title: The large diverse gene family var encodes proteins involved in cytoadherence
 A:Reference number: 220487; MUID:95330813; PMID:7606788
 C:Accession: T28625
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3006 <SUX>
 A:Cross-references: EMBL:LA0609; NID:9886376; PID:9886377; PIDN:AA75397.1
 C:Genetics:
 A:Gene: var-3
 A:Introns: 2597/3

Query Match 31.5%; Score 486; DB 2; Length 3006;
 Best Local Similarity 39.1%; Pred. No. 2.3e-28;
 Matches 115; Conservative 39; Mismatches 108; Indels 32; Gaps 13;

QY 5 KNSNGYPMOCGNINLVDPVCMPPRRKQICVHFLANDNEIKKLOSOVNIKEA 64
 DB 1247 KAGDKQIPGWCNQSQIHTTHNGACMPPRKQICVSGLTFRDRAIR-YYIRFEIKSA 1303
 QY 65 AAEFTFSWYKSKDGEGNELDKELKGIKIPPAFLRSMFTFGDYRDLFGTDI 124
 DB 1304 ALETHFAMDREKEDNGEA---EALKNGNIPDEGKROMYTFGDRDIFGRDIT-HAY 1359
 QY 125 GSKLEQIDSLF-KNGQCK-----SPNGKTRQWTEHSHHEWAMICALV-RIG-AKK 175
 DB 1360 ISGSPVYITILEKENDAKYAAKONSNNELLDWMDQHGDIWEGMICALTHRISDEKK 1419
 QY 176 DDEFTENGYNNVFKSDK-STTLEEFARPOFLRWLEWYDDCYTRQKLYKQVECKRS- 231
 DB 1420 KEIKNKYSYKINKNSPKGSKVDEFAKQFLRWLEWYDDCYTRQKLYKQVECKRS- 231
 QY 232 KSNQDLKCDTECNKCC-----EDYKYYKMKKKEMIPQDKYKDER--DKRFE 276
 DB 1480 KDYDGCK-NTKSNASCVSACKVEYDI--TKKVEYTKQKGNKDAKRTIDREY 1530

RESULT 5

C71625

variant-specific surface protein 1 homolog PFB0010W - malaria parasite (Plasmodium fa
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: C71625
 R:Gardner, M.J.; Tetelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743; PMID:9804551
 C:Accession: C71625
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown.
 A:Molecule type: DNA
 A:Residues: 1-1711 <GAR>
 A:Cross-references: GB:AE001366; GB:AE001362; NID:93845070; PIDN:AA71792.1; PID:9384
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0010W

Query Match 25.7%; Score 396; DB 2; Length 2212;
Best Local Similarity 27.5%; Pred. No. 1e-21;
Matches 101; Conservative 50; Mismatches 94; Indels 122; Gaps 12;

27 VCPMPRRLCLV---HFLANDNEIKKLSQV-----NLEKAFIKSAAE 67
Db 1342 ICIPPRRLRLYOKLHDMASGVYVSGAQTPQGGTSSPSGKEPESDKLRATFLOSAIE 1401
QY 68 TFFSWYVYKSKDGEENLDELKE-
Db 1402 TFFLMDRYKKE---KEIEKKKKYANGVPLSLNGGPPQPGVGTGDSPOSKLOOTGYIP 1457
QY 96 PAFLSMFTEDYRDLFGD---ISKHGE-----GSKLKEQIDSLFKNGDOKSPNG 146
Db 1458 PFFLQMFYTTGADYADIFGKNDYIDTKNGDKDIAEREKKIKDAIEVYKLNMSQPSD 1517
QY 147 KTRQEMWTEHSHETEMALCALV-----KIGAK-----KDDFTENTGYN 185
Db 1518 EKROTWMQNGEHLWNGICALTYKREDEKTPKQNEGLKSALMDKRNKPKDKOXYD 1577
QY 186 NVKFSDS-----TLEEFARPOFLMTEWYDYCYTRQKYLKDVOEKCS 233
Db 1578 KVLIDENSGISPKTNDHVPPLTNFISRPYFYLEWGETTCOREKKLEKIKYECMD 1637
QY 234 ND--OLKCD-----TECNKCEYVYMKKK-EMIPDKRYK 268
Db 1638 EDGKKKCSGDECEIRKODSTVADFYCPCGKYCRPKYMKIEKKKDEYDKOKEAVN 1697
QY 269 DERKKR 275
Db 1698 NQKTAR 1704

RESULT 9

T14029
Variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
C:Accession: T14029
R:Chen, Q.; Barragan, A.; Fernandez, V.; Sundstrom, A.; Schlichtherle, M.; Sahlen, A.; J. Exp. Med. 187, 15-23, 1998
A:Title: Identification of Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1)
A:Reference number: 217860; MUID:96080592; PMID:9419207
A:Accession: T14029
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2228 <CHE>
A:Cross-references: EMBL:AF003473; NID:92961467; PID:92961468; PIDN:AA05730.1
A:Note: FCR3SL.2-vari

Query Match 25.3%; Score 390.5; DB 2; Length 2228;
Best Local Similarity 28.3%; Pred. No. 2.7e-21;
Matches 113; Conservative 48; Mismatches 103; Indels 135; Gaps 18;

8 SNGYDMOC---GGINVEPR--VCPMPRRLCLVHFL-----ANDNEIKKL 50
Db 911 NNSRLGMRCTVPSGPTSSDKNGAICVPPRRRLIYIKIYDNTKTESPOAGSSASST 970
QY 51 -----OSQVNLKFAFIKSAAEFFSWYVYK-----SKDEGNL 85
Db 971 SGTTPPSKALAKAVESAITFFLHMRYKREKAAVQEGAGHGLPVEGSPEDYD 1030
QY 86 DKEIKKIPAPILRSFYFGDRDLFG-----TDISK-----GHGE 124
Db 1031 EDKIKKIPIDGFLKMFYTLGDRDLFGSDNTSVSDTSSSDNKNIVLASSS 1090
QY 125 GSKLKEQIDSL--FRN-----GDOKSPNGKTR--QEWVTEHSHETEMALCALV--KIG- 172
Db 1091 TEGDEREMNNKYEIKNFKRCKSTERSAPNLVSHQPTWMENNGKTYIMHGWCAALTSKDKIAK 1150

QY 173 --AKKDFTEN-----YGINNVKFSKDS-----TILE 197
Db 1151 GVEKKPKIKINPENLMDANKKRPPOYOYNTVTKLIDENSGTSPRTQTQASSDNPTTIL 1210
QY 198 EPARKPOFLMTEWYDYCYTRQKYLKDVOEKCK-----SNDLAKCPT----- 241
Db 1211 HFVRRPYFFRFEWGESFCREKRLKQIKYDKCKENDVGRCSGDGACDSTHDS 1270
QY 242 -----BCNKKCEYVYMKKK-EMIPDKRYK 271
Db 1271 TVSPFPCGCKKCSYRKWKIEFKIEFKQSNAYGOQK 1309

RESULT 10

T28626
Variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28626
R:Su X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrifeldt, J.A.; Peterson, Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: 220487; MUID:95330813; PMID:7606788
A:Accession: T28626
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2664 <SUX>
A:Cross-references: EMBL:LA0609; NID:9886376; PID:9886378; PIDN:AA75398.1
A:Genetics:
A:introns: 2197/3
A:Note: var-2

Query Match 25.0%; Score 385.5; DB 2; Length 2664;
Best Local Similarity 27.4%; Pred. No. 7.9e-21;
Matches 118; Conservative 44; Mismatches 107; Indels 161; Gaps 17;

QY 2 CHPKNSNGY--PDQCGNINLV-----EDPRVCPMPRRLCLVHFL--ANDN 45
Db 1313 CIPKTSNDVTGEGGQ-GSRVYRSTPESGNSDKNGACIIPRRRLVYGLQWANKH 1371
QY 46 EIKKLSQ-----VNLKEAFIKSAAEFFSWYVYK- 76
Db 1372 NTFYSVQGEATRGSEAPAPGKSSSGKREPSDKLRTAFIESAAVETFLMDRYK 1431
QY 77 -----SKOEGNELDELKEKGIAPILRSFYFGDRDLFG--G 115
Db 1432 EWLAKKAELONGGLDYSSGDSDPDPONKLLNGVLPDLRLMFYTLGDRDLVHCG 1491
QY 116 TDSKGHGSS-----KIKQIDSLF-KNGDQ--KSPNGKTRQEWMT 154
Db 1492 NTSQSGNTGSSNNNIVLEASGNKEEDMKIQEKIDQILPKNGTGLVPRKSSAQTPDKMWN 1551
QY 155 EHSHEIWEALCALV-----KIGAKKD----- 177
Db 1552 EHAESIWKMICALITYTEKNPDTSGARGDENKI--EKDEVEYEKEFGSTADKHGTASTPTG 1609
QY 178 -FTENYGINNVKFSKDS-----TLEEFARPOFLMTEWYDYCYTRQKYLK 225
Db 1610 TYKTYDYDEKXVLEPTSGAKTPSASSDPLSDPFLRPYFYLEWQONCKEKRRRLK 1669
QY 226 DVOEKCK-----SNDLKD-----TECNKCEYVYMKK-KKEWIPQ 263
Db 1670 QIKKCEMGSDKRTYSGDEQCDRDRTSNESVADLEGRSCGNSCFYKWKIKRKKREYDKQ 1729
QY 264 DKYYDERDK 273
Db 1730 ANAYSKOKTK 1739

RESULT 11

T28634
Variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum

Db 1154 LDDTSANTGSPVSGEIKTLPDIFSRPFYRIEMGTEFCKERRKRLKIEKRG 1213
QY 234 -----NDQ---LKCDTECNKCEYVYKMK-KKKEMIPQDKY 267
Db 1214 DRTGHEHSGGCGYCTRTDADRNDFVNLNC-RDCHICQCKRKRKWDIKFDEIHKOEKKY 1272
QY 268 KDERDK 273
Db 1273 QGEYDK 1278

RESULT 14

B71600

variant-specific surface protein 1 homolog PFBI055c - malaria parasite (Plasmodium falciparum)

N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: B71600

R:Gardner, M.O.; Yoteliu, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;

Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.

Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

File: Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: B71600

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2197 <GAR>

A:Cross-references: GB:AE001434; GB:AE001362; NID:93845341; PIDN:AC71996.1; PID:9384534

C:Genetics:

A:Experimental source: clone 3b7

A:Gene: PFBI055c

Query Match 22.0%; Score 339.5; DB 2; Length 2197;
Best Local Similarity 25.2%; Pred. No. 1.9e-17;
Matches 107; Conservative 50; Mismatches 114; Indels 153; Gaps 17;

QY 2 CHPKNSNG---YPMOC---GNINLV-----EDPRVCMPPRQKICVHFL-- 41
Db 931 CLKXGPGCKERFPNMKCVSSGKSVATAGSSGATGKSGDKALICVPRRRRLVYGGILK 990
QY 42 -----AND---NEIKKLSQVNLKEAFIKSAAEETPFSWYYSK 78
Db 991 LTSAGTSSSPQGSSESRASDVSOQNGCGDDTTTSTSLKMFETFALETFFLMHRYKE 1050
QY 79 -----DGEENELDKEL-KEGKIPPAFLRSMTFFGDRFL 113
Db 1051 WEAKKAEIQRNGLLGTGASLNLGGDSNPQTOLQKSGTIPDLRLMFLYLGDRIL 1110
114 F-GIDISGHEGEG-----SKLEQJDSLFKNGDOKSPNG-----KTRQEW 152
1111 VRGVAADNKGNNIILNAGNKDEKQKMEIKQEKLEQILPTSGNKETRGPONSVNDROSL 1170
QY 153 WTEHSHEIWEAMLCAL-----VKIGAKKDDFTEN-----YGYNNVK 188
Db 1171 WDRIAEHVWGHGVCALTYKDDNGLKGYAKKPKQIENPEKLMNETTKPKDEKYOYQAK 1230
QY 189 FSDSK-----TLLEFAKRPQFLRLTMYDYCYTRQKYLKDVQEKRSN--- 234
Db 1231 LEDESGEKRPDSASGKTLTDFIKRPFYRLEWGENCKKRTMLGKIKEDCYKNGR 1290
QY 235 ---DQKLC-----DTE-----CNKKCEDYKYMKKKKE-----WIPDKYYK 268
Db 1291 CSGGGLKCNELIVIDKEIFGDLCLPTCARHCFYKRWINTKRDENKOSNAYSEQKKYE 1350
QY 269 DERD 272
Db 1351 EEND 1354

RESULT 15

T18396

erythrocyte membrane protein variant 2 - malaria parasite (Plasmodium falciparum) (fragm
C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18396
R:Baruch, D.I.; Pasloske, B.L.; Singh, H.B.; Bi, X.; Ma, X.C.; Feldman, M.; Taraschi,
Cell 82, 77-87, 1995
A:Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen a
A:Reference number: 218925; MUID:95330812; PMID:7541722
A:Accession: T18396
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1729 <BAR>
A:Cross-references: EMBL:U27339; NID:9914920; PID:9914921; PIDN:AAA89134.1
C:Genetics:
A:Gene: EMP1
A:Note: var-2

Query Match 19.4%; Score 299.5; DB 2; Length 1729;
Best Local Similarity 22.6%; Pred. No. 1.6e-14;
Matches 99; Conservative 47; Mismatches 121; Indels 171; Gaps 15;

QY 2 CHPKNSNGYPMOCGNINLVEDPR-----VCMPPRQKICVHFL----- 41
Db 854 CLKXGKTAPISMKC-----IPSDTKSVATGSDTGSGLICVPRRRRLVYKGLHDMAG 909
QY 42 -----ANDN-----BIKKLSQVNLKEAFIKSAAEETFS 71
Db 910 ETTEAKSOETSOGQKTPSGNESSPSEKLPQGPPTPTETKETSILHAFVSPRLRFLP 969
QY 72 WYYSK-----DGEENELDKELKEGKIPPAFLRSMTFFGDRYDF 112
Db 970 WAKKEQKKAQAGATGLQPLGTVVDSDDPQOTOLKRGNIPLNDFLQMFYTLGDYNDI 1029
QY 113 LEQTD-----ISKHGEGSKLKEQIDSLFKNQDOKSPNGK----- 147
Db 1030 CIGGDRDVGDTIVSITGESYKKIKSILIEGFLKKQYTVSPSPDTSRPFVHPQTSVE 1089
QY 148 -TROEMWTEHSHEIWEAMLCALV-----KLT-----GAK--- 174
Db 1090 KTPQQTWEANGPHLWNGMICALTYEDSGAIGQPPQKVEDADKYLEKLPNTANGIKMYL 1149
QY 175 KDDFTENGYNNVFKSDKST-----TLLEFAKRPQFLRLTMYDYCYTRQKYLKDVQ 228
Db 1150 KEDNTSSAMPTSSSSSSSGSNDPINTPKLTERVEITFFRYLHEWQONCKEKMRKIKOIT 1209
QY 229 EKCKSND-----QKLC-----DT-----ECNKCEDYKYMK 255
Db 1210 KECKVGENGYGRKQKTPQSCGYCEDCEDQSKSYDTVADLECPKCAKHCRWYKKNIE 1269
QY 256 KKK-EMIPQDKYKDERD 272
Db 1270 KKDEFTQEKAFPRQKD 1287

Search completed: April 28, 2003, 10:32:56
Job time : 16.4383 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 2.65193 Seconds

(without alignments)
4316.649 Million cell updates/sec

Title: US-10-087-013-2-COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPEKNSNGTPDMCCGNIN.....KKWIPQDKYKDEPKRRF 276

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	274	17.8	1435	1 EBA1_PLAFC	P19214 plasmodium
2	236	15.3	1153	1 PVD8_PLAFC	P50493 plasmodium
3	223	14.5	1070	1 PVD8_PLAFC	P22290 plasmodium
4	218.5	14.2	1073	1 PVD8_PLAFC	P22345 plasmodium
5	214.5	13.9	1070	1 PVD8_PLAFC	P50494 plasmodium
6	105.5	6.8	346	1 YET7_METUA	P05872 methanococ
7	103.5	6.7	677	1 RPOC_SPIOL	P11705 sphinctra ol
8	99	6.4	3660	1 DMD_CHICK	P11533 gallus gall
9	98.5	6.4	1341	1 RPA1_METUA	P05845 methanococ
10	97.5	6.3	680	1 RPOC_TOBAC	P12116 nicotiana t
11	95.5	6.2	678	1 GARP_PLAFC	P13816 plasmodium
12	95	6.2	4427	1 PKSL_BACSU	P05470 bacillus su
13	94.5	6.1	446	1 TBB1_NOTCO	P36521 notochenia
14	94.5	6.1	1178	1 MNM4_YEAST	P36644 saccharomy
15	94	6.1	1081	1 SYT_TETTH	P36422 tetrahymena
16	92.5	6.0	441	1 Y107_METUA	P06036 methanococ
17	92.5	6.0	450	1 TBB2_CAEEL	P52275 caenorhabdi
18	92	6.0	451	1 TBB2_CAEEL	P34108 naegleria g
19	91	5.9	434	1 TBBQ_HUMAN	P09867 homo sapien
20	91	5.9	444	1 TBB1_HUMAN	P07437 homo sapien
21	91	5.9	444	1 TBB5_HUMAN	P04350 homo sapien
22	91	5.9	444	1 TBB7_CHICK	P09244 gallus gall
23	91	5.9	444	1 TBBX_HUMAN	P05218 homo sapien
24	91	5.9	450	1 TBB_BOMO	P41385 bombyx mori
25	91	5.9	617	1 Y237_BUCAL	P57331 buchera ap
26	91	5.9	1230	1 TBBE_YEAST	P50101 saccharomy
27	90.5	5.9	445	1 TBB_PSEAM	P09140 pseudopur
28	90.5	5.9	446	1 TBB6_CHICK	P09207 gallus gall
29	90	5.8	467	1 TBB1_PHYPO	P07336 physarum po
30	90	5.8	1391	1 MST2_DROHY	P08696 drosophila
31	89.5	5.8	2351	1 FAS_HUMAN	P00451 homo sapien
32	89.5	5.8	680	1 RPOC_ARATH	P56763 arabidopsis
33	89.5	5.8	1164	1 BAC_STRAG	P27951 streptococ

34	89	5.8	444	1 TBB_ONCGI	P41387 onchocerca
35	89	5.8	446	1 TBB_GIALA	P05304 giardia lam
36	89	5.8	712	1 ARS2_CAEEL	P06615 caenorhabdi
37	88.5	5.7	441	1 TBB7_CAEEL	P17299 caenorhabdi
38	88.5	5.7	441	1 TBB7_CAEEL	P12456 caenorhabdi
39	88.5	5.7	449	1 TBB4_CHICK	P06652 gallus gall
40	88.5	5.7	450	1 TBB3_MOUSE	P09487 mus musculu
41	88.5	5.7	450	1 TBB4_HUMAN	P13509 homo sapien
42	88.5	5.7	734	1 YKR2_CAEEL	P34308 caenorhabdi
43	88.5	5.7	944	1 NUF1_YEAST	P32380 saccharomy
44	88	5.7	442	1 TBB_STYLE	P1857 stylonychia
45	88	5.7	448	1 TBB1_BRUPA	P18241 brugia paha

ALIGNMENTS

RESULT 1	ID	EBAL_PLAFC	STANDARD:	PRT: 1435 AA.
AC	P19214	EBAL_PLAFC	01-NOV-1990 (Rel. 16, Created)	
DT	01-FEB-1996 (Rel. 33, Last sequence update)			
DT	01-FEB-1996 (Rel. 33, Last annotation update)			
DE	Erythrocyte-binding antigen EBA-1/5.			
OS	Plasmodium falciparum (Isolate Camp / Malaysia).			
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.			
OX	NCBI_TaxID=5835;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=90377299; PubMed=2204835;			
RA	Slm B.K.L.;			
RT	*Sequence conservation of a functional domain of erythrocyte binding			
RL	antigen 175 in Plasmodium falciparum.*			
CC	Mol. Biochem. Parasitol. 41:293-296(1990).			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement. (See http://www.isb.ch/announce/			
CC	or send an email to license@isb.slb.ch).			
DR	EMBL; X52524; CAA36756.1; -			
DR	PIR; S11561; S11561.			
KW	Antigen.			
FT	DOMAIN 159 1104			
FT	ESSENTIAL FOR BINDING TO			
FT	ERYTHROCYTES.			
FT	E -> V (IN STRAINS FCR-3 AND ITG).			
SQ	VARIANT 1031 1031			
SEQ	SEQUENCE 1435 AA; 167389 MW; 3244309021BIC3D6 CRC64;			
Query Match	17.8%; Score 274; DB 1; Length 1435;			
Best Local Similarity	29.2%; Pred. No. 1.3e-13;			
Matches	76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;			
QY	14 WOCGNTLVDPVCMPPRRQKLCVHPLANDNEIKLQSOVNLKFAFKSAAEFFSWY 73			
DB	474 WECKNPILISTKOVCPRRQELC--LGNIDRIYD-KNLMLEKEHLALAIATESRLKR 529			
QY	74 YYSKSGEGNELKEKGIKIPAFILSMFYTGDFLFGDIDSGHEGSKLKQID 133			
DB	530 KYRNKD-----DREV-----CKIINKTFADIDIDIGTDYVNDL-SNRKLVGKIN 573			
QY	134 SLKNGDKSPNGKT-ROEWMTESHSHIWMALCALVIAKKDDPLENGYNNVFSDK 192			
DB	574 TNSKTYHRRKNKDLFDEWKKYIKKDVNVT-----SWFKDK 612			
QY	193 STLEEFPAKR-POFLRWLTEDYDYCYTRQRYKLDVQEKSKNDOLKC-DTECNKCEDY 250			
DB	613 TVCKEDDIENIPQFRRFSSWGDYCDQDKTKMETLVECK---EKRCEDDNCKSKNSY 669			
QY	251 VKYM-KKKKEMIPQDKYKD 269			

Db 670 KEMISKKEEYKAKOYOE 689

RESULT 2
PVDR_PLAKN STANDARD; PRT; 1153 AA.

AC P50433;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, beta form precursor (Erythrocyte binding protein).
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;

SEQUENCE FROM N.A.
MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Slim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.,
Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
-1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
DETERMINANT.

-1- SUBCELLULAR LOCATION: Type I membrane protein.
-1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M90694; AAA29603.1;
Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
Multi-gene family.

FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 1153 DUFFY RECEPTOR, BETA FORM.
FT DOMAIN 22 1085 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1107 1153 POTENTIAL.
FT DOMAIN 1107 1153 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 179 179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 202 202 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 348 348 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 467 467 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 626 626 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 722 722 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 847 847 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 856 856 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 900 900 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 910 910 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 935 935 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1153 AA; 130471 MW; 6497BD16CE7BEC01 CRC64;

Query Match 15.3%; Score 236; DB 1; Length 1153;
Best Local Similarity 27.2%; Pred. No. 9.2e-11;
Matches 74; Conservative 45; Mismatches 95; Indels 58; Gaps 13;

Db 6 KNSNGYDPOGGININLVDPVCMPPRRQKLCVHFAN-----DNEIKLQSOVNL 56
217 KRKGSTDMOCPT-----KKVCIPIRRYQCMKELNINVNNTKTHSHNITEL--KLNL 269
OY KEAFIKSAAEFTFSWYYSKSGDEGNEDEKELKEGKIIPAFILSMFYTGDRDPLFEGT 116
Db 270 KEKITVDAAEVGGILLKKYNNVSE--DICKDKI-----WSIEDRGDILMG 314
OY 117 DISKNGEGSKLEQIDSLFKNGDQKSPNGKTRQEMWTEHSHEIWEAMCALVYKIGAKKD 176

Db 315 DM-EGIGGYOVENNIRIVFGTGTQTD---RKKWNESKRYIWEATLLSVKK----- 364

OY 177 DFTENGYN--NVKESDKSTLEEFARQPLWLEWYDQCYTRQKYLKQVSKCK-- 232
Db 365 ---KLNGYSAMNCK-EDVQINVE-----PQIRWIREMNDYMSLEPQRIKREKCDRK 415

OY 233 ---SNDOLKCDUECNKCKEDYKYM-KKKKEW 260
Db 416 LYTNIRICTMSPCNSCKLYDOWITRRKKW 447

RESULT 3

PVDR_PLAYS STANDARD; PRT; 1070 AA.

AC P22290;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Duffy receptor precursor (Erythrocyte binding protein).
OS Plasmodium vivax (strain Salvador I).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=126793;

SEQUENCE FROM N.A.
MEDLINE=91187056; PubMed=1489231;
RA Fang X., Kaslow D.C., Adams J.H., Miller L.H.,
Cloning of the Plasmodium vivax Duffy receptor.*;
Mol. Biochem. Parasitol. 44:125-132(1991).
-1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
DETERMINANT.

-1- SUBCELLULAR LOCATION: Type I membrane protein.
-1- SIMILARITY: HIGH, TO P.KNOWLESI DUFFY RECEPTORS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: M61095; AAA63423.1;
Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1070 DUFFY RECEPTOR.
FT DOMAIN 21 1007 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1008 1025 POTENTIAL.
FT DOMAIN 1026 1070 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 420 420 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 715 715 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 767 767 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 825 825 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 903 903 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 938 938 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1070 AA; 119683 MW; CB051DF13E294603 CRC64;

Query Match 14.5%; Score 223; DB 1; Length 1070;
Best Local Similarity 26.1%; Pred. No. 8.6e-10;
Matches 70; Conservative 44; Mismatches 90; Indels 64; Gaps 14;

OY 13 DMQCGININLVDPVCMPPRRQKLCVHFAN-----DNEIKLQSOVNLKEAFI 61
Db 227 DMDC---NTRKD--VCIPDRRYQCMKELNINVNNTDFHNDIIFPKLY---LKRKIT 277
OY 62 KSAAEFTFSWYYSKSGDEGNEDEKELKEGKIIPAFILSMFYTGDRDPLFEGTDISKG 121
Db 278 YDAAYE-----GDLILKLNRYKNDPFCIDIMWSLGRDILMGTD--EG 321
OY 122 HEGSKLKEQIDSLFKNGDQKSPNGKTRQEMWTEHSHEIWEAMCALVYKIGAKKDPTFN 181

```

Db 322 IGSKVENVNLSIF-GTDERAO--ORRKNWNESKAQITAMYSVK--RLKGNFTWT 376
Oy 182 YGVN---NWKFSKSTLEEFARPOFLNLTWMDYDCTYRQKLDVQEKSS---N 234
Db 377 CRNNAVNNE-----POLYRIRRMGRDIYSELFTEYQKLKREKDGKINTY 422
Oy 235 DQKCDT-ECKKCEDYVKYM-KKKKEW 260
Db 423 DKRVCKVPCONACKSYDQWITRRKKQW 450

RESULT 4
PVDG_PLAKN STANDARD; PRT; 1073 AA.
ID PVDG_PLAKN STANDARD; PRT; 1073 AA.
AC P22545;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, alpha form precursor (Erythrocyte binding protein).
Plasmodium knowlesi.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5850;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.,
RA Adams J.H., Hudson D.E., Toril M., Ward G.E., Wellens T.E.,
RA Aikawa M., Miller L.H.;
RT "The Duffy receptor family of Plasmodium knowlesi is located within
RT the micronemes of invasive malaria merozoites.";
RL Cell 63:141-153(1990).
CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
CC DETERMINANT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; M90466; AAA29602.1; -
DR EMBL; M68517; AAA29590.1; -
DR EMBL; M68518; AAA29591.1; -
DR PIR; A35970; A35970.
KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
KW Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1073
FT DOMAIN 22 1007
FT TRANSMEM 1008 1029
FT DOMAIN 1030 1073
FT CARBOHYD 134 134
FT CARBOHYD 179 179
FT CARBOHYD 202 202
FT CARBOHYD 252 252
FT CARBOHYD 348 348
FT CARBOHYD 679 679
FT CARBOHYD 746 746
FT CARBOHYD 779 779
FT CARBOHYD 788 788
FT SEQUENCE 1073 AA; 120683 MW; 3965FC9F46B71808 CRC64;
Query Match 14.2%; Score 218.5; DB 1; Length 1073;

```

```

Best Local Similarity 25.9%; Pred. No. 1.9e-09;
Matches 71; Conservative 47; Mismatches 101; Indels 55; Gaps 15;
Oy 2 CHEPKNSNGYPMQCCGINILVEDPRVCMPPRRKLCVHFLAN-----DNEIKKLS 52
Db 214 CNDKR-KRGRRMDC-----PAKDCICISVRQOLCKGLTNLVNRTTSHNDITFL-- 265
Oy 53 QVNLKEAFIKSAAAEFTFSWYKKSKDGEENELDKELKEGKIPPAFLRSMEFYFGDYDF 112
Db 266 KLNLRKRLMYDAV-----EGDLLLRK-NNYQYNKEFCQKIDRWIGDGDGI 310
Oy 113 LFGTDSKSGHSGSKLEQDLSLFKNGDOKSPNGKTRQEWTEHSHETWAMCALVKIG 172
Db 311 IMGTNM-EGIGYSQVYENNROYF-GTDERAK--QDRKQWNNSKEHIMAMFESI----- 362
Oy 173 AKKDDFENYNNVNFSDKSTLEEFARPOFLNLTWMDYDCTYRQKLDVQEK 232
Db 363 --KSLRKEF-----VWCKRKDYLR--VEPOLYRIRRMGRDIYSELFTEYQKLKREK 413
Oy 233 S---NDQKCDT-ECKKCEDYVKYM-KKKKEW 260
Db 414 SKLYNNMAICMLPLCHDACKSYDQWITRRKKQW 447

RESULT 5
PVDG_PLAKN STANDARD; PRT; 1070 AA.
ID PVDG_PLAKN STANDARD; PRT; 1070 AA.
AC P50494;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, gamma form precursor (Erythrocyte binding protein).
Plasmodium knowlesi.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5850;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.;
RA Adams J.H., Hudson D.E., Toril M., Ward G.E., Wellens T.E.,
RA Aikawa M., Miller L.H.;
RT "A family of erythrocyte binding proteins of malaria parasites.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
CC DETERMINANT.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; M90695; AAA29604.1; -
KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
KW Multigene family.
FT SIGNAL 1 21
FT CHAIN 22 1070
FT DOMAIN 22 1003
FT TRANSMEM 1004 1025
FT DOMAIN 1026 1070
FT CARBOHYD 134 134
FT CARBOHYD 179 179
FT CARBOHYD 676 676
FT CARBOHYD 743 743
FT CARBOHYD 785 785
FT CARBOHYD 936 936
FT SITE 279 281
FT SEQUENCE 1070 AA; 120931 MW; 703D68811BC11B50 CRC64;
Query Match 13.9%; Score 214.5; DB 1; Length 1070;
Best Local Similarity 23.9%; Pred. No. 3.9e-09;

```

Matches 68; Conservative 48; Mismatches 91; Indels 77; Gaps 14;

QY 2 CHPKNSNGYPMOCNINIVEDPVCMPPRROKLCVHFLAN-----DNEIKLQ 51
 Db 214 CNDRK-KRGEDMDCPT-----ENDVCIPDRRIOLCMETITNLYDDTHFHSIDIFKRSY 267
 QY 52 SQVNL-----KRAFKSAAEETFFSWYKSKDGBNELDEKCKIPPAFLRSMFY 104
 Db 268 SRRLIYDVGGRGDLTK-----KYNMYSEDLCKDK-----W 301
 QY 105 TFGDPRDFLEGTIDSKGHSKLEQIDSLFNKGOKSNGKTRQDQWTEHSHEIWEAM 164
 Db 302 SLDPFGDIIINGTDM-BEIGSLVYVANNKSIPTGTSAEID---RKMMWDRKDKIWKAM 357
 QY 165 LCLAVKIGAKKDDFTENYGVNKKFSDKSTTLEFAKRPQFLMTEWYDDYCYTROXYL 224
 Db 358 ILSVKE-----KNYSAMNCK-EDVQIKVE-----PQIYMIEMNGMDYASERREOR 403
 QY 225 KDVOEKCKSNOLKCDT-----ECNKCEYVVKYMK-KKKEM 260
 Db 404 RKLNEKE--DKLYSTMLICTLPCNNACKSYDEWITGKKOW 445

RESULT 6
YE77_METJA

AC 058872; STANDARD; PRT; 346 AA.

DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein MJ1477.

GN MJ1477.

OS Methanococcus jannaschii.
 OC Archaea: Euryarchaeota: Methanococci: Methanococcales;
 CC Methanocaldococcaceae; Methanocaldococcus.

OK NCBI_TaxID=2190;

RP SEQUENCE FROM N.A.

RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=6868087;

RA Bult C.J., White O., Olsen G.V., Zhou L., Fleischmann R.D.,
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Claydon R.A., Gocayne J.D.,
 RA Kervaege A.R., Dougherty B.A., Tomb J.F., Adams M.D., Reisch C.L.,
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
 RA Uitterlinden T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 RA Blank H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 NL Science 273:1058-1073(1996).

CC -1- SIMILARITY: TO D.RADIODURANS DR0705 AND T.MARITIMA TM1410.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: U67588; AAB99489.1; -

DR DR TIGR: MJ1477; -

DR InterPro: IPR004352; TM1410-related.

DR Pfam: PF03537; DUF297.1.

DR PRINTS: PR01545; THEMAYE10DUF.

KM Hypothetical protein; Transmembrane; Complete proteome.

FT TRANSMEM 16 36 POTENTIAL.

SO SEQUENCE 346 AA; 40620 MW; A71B3D69A956BCE CRC64;

Query Match 6.8%; Score 105.5; DB 1; Length 346;
 Best Local Similarity 22.1%; Pred. No. 0.31; Mismatches 93; Indels 14;

QY 43 NDNEIKIOSQVNLKRAFKSAAEETFFSWYKSKDGBNELDEKCKIPPAFLRSMFY 96
 Db 63 NSNNLKFLMA-YOLQADIDELIANSFTLIYIDYSGDGENKRSYSEELKTKKAKTPT 121
 QY 97 AFLRSFETFGDYRDLFTGTD-----ISKHGECSK-LKDOI 132
 Db 122 AIT-----SIGEAEDRFYWDNEMLKNPPKWLGDENPEWEGCYAVKVMHPEWKILFSYL 176
 QY 133 DSLFNKGDKSPNGKTRQ-EMWTEHSHEIWEAMLCALYKIGAKKDDT-----E 180
 Db 177 DKIIQGGFCGYVLDVYDEEYEAENGID-----EDFYAKEMIKFYEIS 220
 QY 181 NGYNNVKS-----DKSTLEFAKRPQFLMTEW-YDYCY--TRQXYLKVOEKCK 232
 Db 221 NCCRNKTNNSFTIIPONGERLEYDKHGLNMTVASGMAVEDLFYDGVGQKTEEELNERIK 280
 QY 233 SNOQLKCD-----TECN--KKCD-----YKYMCK 256
 Db 281 LLDKVKDSGKFLVYDDYDDGKTINENIKRVEDFINKSLDKGYVPYAK 329

RESULT 7
RPOC_SPIO

AC P11705; STANDARD; PRT; 677 AA.

DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6).

GN Spinacia oleracea (Spinach).

OC Chloroplast
 CC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;
 CC Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots;
 CC Caryophyllales; Caryophyllales; Chenopodiaceae; Spinacia.

OK NCBI_TaxID=3562;

RP SEQUENCE FROM N.A.

RC STRAIN-cv. Giant d'hyver, and cv. Monato;

RX MEDLINE=21187424; PubMed=11292076;

RA Schmitz-Limweder C., Maier R.M., Alcaraz J.-P., Cottet A.,
 RA Hermann R.G., Mache R.;
 RA "The plastid chromosome of Spinacia (Spinacia oleracea): complete
 RA nucleotide sequence and gene organization.";
 RL Plant Mol. Biol. 45:307-315(2001)

CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.

CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC (RNA)(n).

CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
 CC SUBUNITS: ALPHA, BETA, BETA', AND BETA".

CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AJ400848; CAB88716.1; -

DR DR PIR: B29959; B29959.

DR HSSP: Q9KRW6; 1HOM.
DR InterPro: IPR000722; RNA_pol_A.
DR Pfam: PF00623; RNA_pol_A.1.
KW Transference; Transcription; DNA-directed RNA polymerase; Chloroplast.
SEQUENCE 677 AA: 78149 MW: D618E17A829CEBD CRC64;
Query Match 6.78; Score 103.5; DB 1; Length 677;
Best Local Similarity 24.68; Pred. No. 0.95;
Matches 51; Conservative 22; Mismatches 53; Indels 81; Gaps 13;
QY 16 CGNINLV-----EDPRVCMPPRROKLCVHFLANNNEIKLOSYNLKAFFKSAAEFFS 71
DB 71 CGNRYVIGDEKEDPKRC-----EGCGVEFY--DSRIKRYGM-----GYIKLACPVTHV- 116
QY 72 WYVYKSKDGEENDELK--EGKI-----PPAFLR-----S 101
DB 117 WYVYKSKDGEENDELK--EGKI-----PPAFLR-----S 101
QY 102 MFTY---FGDYRDFLFGTDSKSGHGEKSKKEQ-----IDSLFKNDQKSPNCKTR 149
DB 177 LFTTGGEDTFRN---REISTGAG--AIREQDLADLRLTIIDSPFAWEKELGEGSTG 229
150 QEWTEHSEHWEMCALVKIGAKKD 176
DB 230 NEMEDR-----KVGRRKD 242
RESULT 8
DMD_CHICK
ID DMD_CHICK STANDARD; PRT; 3660 AA.
AC P11533;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dystrophin.
GN DMD.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89098331; PubMed=3062582;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "Nucleotide sequence of chicken dystrophin cDNA."
RL Nucleic Acids Res. 16:11815-11815(1988).
RN [2]
SEQUENCE FROM N.A.
TISSUE=Muscle;
MEDLINE=89210800; PubMed=3072195;
RA Lemaire C., Heilig R., Mandel J.L.;
RT "The chicken dystrophin cDNA: striking conservation of the C-terminal
coding and 3' untranslated regions between man and chicken."
RL EMBO J. 7:4157-4162(1988).
CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
CC PLASMA MEMBRANE.
CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
CC ABP-120, ABP-180, OR BETA-FODRIN).
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
CC -1- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: X13369; CAA31746.1; -.
DR PIR: S02041; S02041.
DR HSSP: P46939; 1BHD.
DR InterPro: IPR001589; Actbind.actin.
DR InterPro: IPR001715; Calponin-like.
DR InterPro: IPR002017; Spectrin.
DR InterPro: IPR002349; WW.
DR InterPro: IPR001202; WW_Rsp5_WWP.
DR InterPro: IPR000433; Znf_Z2.
DR Pfam: PF00307; CH; 2.
DR Pfam: PF00397; WW; 1.
DR Pfam: PF00435; spectrin; 22.
DR Pfam: PF00569; Z2; 1.
DR PRINTS: PR00403; WMDOMAIN.
DR SMART: SM00033; CH; 2.
DR SMART: SM00150; SPEC; 21.
DR SMART: SM00456; WW; 1.
DR SMART: SM00291; Znf_Z2; 1.
DR PROSITE: PS00019; ACTININ_1; 1.
DR PROSITE: PS00020; ACTININ_2; 1.
DR PROSITE: PS50021; CH; 2.
DR PROSITE: PS01159; WW_DOMAIN_1; 1.
DR PROSITE: PS50020; WW_DOMAIN_2; 1.
DR PROSITE: PS01357; ZF_Z2_1; 1.
DR PROSITE: PS50135; ZF_Z2_2; 1.
DR Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; Zinc-finger.
FT DOMAIN 1 244 ACTIN-BINDING.
FT DOMAIN 19 123 CH 1.
FT DOMAIN 138 241 CH 2.
FT REPEAT 341 449 SPECTRIN 1.
FT REPEAT 450 558 SPECTRIN 2.
FT REPEAT 561 669 SPECTRIN 3.
FT REPEAT 721 830 SPECTRIN 4.
FT REPEAT 832 936 SPECTRIN 5.
FT REPEAT 945 1047 SPECTRIN 6.
FT REPEAT 1050 1156 SPECTRIN 7.
FT REPEAT 1159 1265 SPECTRIN 8.
FT REPEAT 1268 1369 SPECTRIN 9.
FT REPEAT 1470 1570 SPECTRIN 10.
FT REPEAT 1573 1678 SPECTRIN 11.
FT REPEAT 1681 1782 SPECTRIN 12.
FT REPEAT 1879 1981 SPECTRIN 13.
FT REPEAT 2013 2103 SPECTRIN 14.
FT REPEAT 2106 2211 SPECTRIN 15.
FT REPEAT 2214 2321 SPECTRIN 16.
FT REPEAT 2472 2574 SPECTRIN 17.
FT REPEAT 2577 2683 SPECTRIN 18.
FT REPEAT 2686 2799 SPECTRIN 19.
FT REPEAT 2802 2904 SPECTRIN 20.
FT REPEAT 2906 2928 SPECTRIN 21.
FT REPEAT 2931 3037 SPECTRIN 22.
FT DOMAIN 3052 3085 WW.
FT ZN_FING 3304 3351 Z2-TYPE.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1171 1171 MISSING.
FT VARIANT 1869 1869 Q -> H.
FT VARIANT 1885 1885 K -> R.
SQ SEQUENCE 3660 AA: 422874 MW: 85493DAF6D5B6DA CRC64;
Query Match 6.48; Score 99; DB 1; Length 3660;
Best Local Similarity 18.78; Pred. No. 15;
Matches 53; Conservative 49; Mismatches 90; Indels 92; Gaps 11;
QY 45 NEIKLQSOYNLKEAFKSAAEFFFSMYYSKDGGENL-----DKELEKGIIPAF 99
DB 1861 NALKDLRSQRKK-----ALEISHQYQYKR---QADDLMTYLDIDIEKKIASLP----- 1906
QY 100 RSMFTYFGDYRDFLFGTDSKSGHGEKSKLKEQDLSLFKNDQKSPNCKTRQEWTEHSE 159
DB 1907 -----DHKD---EQKLKEIGELKEKKEDLVANVRQAEKRLSKDGAAR----- 1945
QY 160 IWEAMLCALVKIGAKKDPTFENYG--YNNVFSKDSSTLLE----- 198

Db 1946 ---AABPTLVQLSKRWDESKFAQFRRLNVAIOVLDTTFTVMTESMTVETVPSSTY 2002
 QY 199 FAKRPFLEMLTEWYDYCYTRQKYLKDVOECKSNDLQKOTECNKCEYVYKMKKK 258
 Db 2003 LAEIIQLQALSE-VEERLNSPYLAKOCEDLLKOEELCKNIKIDGLGRIQIHDIIHSKK 2061
 QY 259 -----EWIPDKYKXDER---DKKR 275
 Db 2062 TPALASAPRETANTIODKLTOINSQWEKYNKMYRDRQARFPRKSK 2105
 RESULT 9
 ID RPA1_METUA STANDARD; PRT; 1341 AA.
 AC 058445;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6) [Contains: Mja
 RPA1 OR MJ1042.
 Methanococcus jannaschii.
 Archaea; Euryarchaeota; Methanococci; Methanococcales;
 Methanocaldococcaceae; Methanocaldococcus.
 NCBI_TaxID=2190;
 [1]
 SEQUENCE FROM N.A.
 STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
 RX MEDLINE=96337999; PubMed=8688087;
 RA Bilt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
 Sulten G.G., Blake O., Olsen G.J., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
 Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
 Scott J.L., Geochagen N.S.M., Weidman J.F., Fuhman J.L., Nguyen D.,
 Uffebach T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
 Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
 Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
 "Complete genome sequence of the methanogenic archaeon, Methanococcus
 jannaschii.";
 Science 273:1058-1073(1996).
 RL Science 273:1058-1073(1996).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA}(N).
 CC -1- COFACTOR: ZINC.
 CC -1- PTM: THIS PROTEIN UNDERGOES A PROTEIN SELF SPLICING THAT INVOLVES
 A POST-TRANSCRIPTIONAL EXCISION OF THE INTERVENING REGION (INTEIN)
 FOLLOWED BY PEPTIDE LIGATION (POTENTIAL).
 CC -1- SIMILARITY: THE COMBINED A+A' SUBUNITS CORRESPOND TO THE A
 SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
 EUKARYOTIC BETA' SUBUNIT.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation
 at the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC DR EMBL: U67547; AAB99046.1;
 DR TIGR: MJ1042;
 DR InterPro: IPR003586; Hedgehog_hntc.
 DR InterPro: IPR003587; Hedgehog_hntn.
 DR InterPro: IPR002203; Intein.
 DR InterPro: IPR004042; Intein_endonc.
 DR Pfam: PF00623; RNA_pol_A_1.
 DR SMART: SM00305; Hntc; 1.
 DR SMART: SM00306; Hntn; 1.
 DR PROSITE: PS50818; INTEIN_C_TER; 1.
 DR PROSITE: PS50819; INTEIN_ENDONUCLEASE; 1.

DR PROSITE: PS50817; INTEIN_N_TER; 1.
 KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
 KW Autocatalytic cleavage; Protein splicing; Complete proteome;
 KW Zinc-finger.
 FT CHAIN 1 460
 FT CHAIN 911
 FT CHAIN 912 1341
 FT ZN-FLNG 62 105
 FT SEQUENCE 1341 AA; 152780 MW; 4D98546ED552A12 CRC64;
 Query Match 6.4%; Score 98.5; DB 1; Length 1341;
 Best Local Similarity 19.8%; Pred. No. 5.1;
 Matches 71; Conservative 51; Mismatches 111; Indels 125; Gaps 16;
 QY 3 HPRKNSGYDPMQCNINIVEDPRVCMPPRRQKLCVHPLANDNEIKLQSOVNLKEAFIK 62
 Db 533 HPRYTNGKR--RCGBLKVGVDEVIITP-----NDFPFEDNRYVDEEKIK 577
 QY 63 SAALTFESWYIYKSKDGNELKELKGIIPAFRLSMFYEGYDPLFGTDSKGH 122
 Db 578 KYINNGT---YK-----NKIINELDKRLIP-----LYNDQKASILARIYGHV 621
 QY 123 GEGS-----KLKQIDSLFNGDQ-----KSPNKTQEW 152
 Db 622 GDGSLINNNKNSRVYFRGDIEDLTKIKEDLKLGYDSEIKLHGEEITDYNCKRKRIK 681
 QY 153 WHEHHEIWEALCALVR-IGAKKDFTEN-YGVNN----- 186
 Db 682 GKGYEFYRKSLICILKALCGVGDKTKMYGIPNMTKTAPKYIKKEFLSAYGSELT 741
 QY 187 VFSKSTLTLEFAKRPQELMLTEWYDYCYTRQKYLKDVOECKS-NDLQKOTECNK 245
 Db 742 PRIRHNGSFELEK---IAKIEIIDE-----DREIKDKELKFGLEIKRYVEGN 793
 QY 246 -----KEDIYKYKKKKKKW-----IPDKYKXDERDK 274
 Db 794 LRKGYKTKYVVASIYNHKEFFGRIGTYANKKFTLARVAYEVLIMKELIKDRNIK 851
 RESULT 10
 ID RPOC_TOBAC STANDARD; PRT; 680 AA.
 AC P12116;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA-directed RNA polymerase beta' chain (EC 2.7.7.6).
 GN RPOC1.
 OS Nicotiana tabacum (Common tobacco).
 CC Chloroplast.
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 CC Asteridae; euasterids I; Solanales; Solanaceae; Nicotiana.
 CC NCBI_TaxID=4097;
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-cv. Bright yellow 4;
 RA Shinozaki K., Ohme K., Tanaka M., Wakasugi T., Hayashida N.,
 Matsubayashi T., Zaita N., Chunwongse J., Obokeata J.,
 Yamaguchi-Shinozaki T., Ohno C., Torazawa K., Meng B.-Y., Sugita M.,
 Tohoh N., Shimada H., Sugitara M.;
 RA Deno H., Kamogashira T., Yamada K., Kusuda J., Takaiwa F., Kato A.,
 RT "The complete nucleotide sequence of the tobacco chloroplast genome:
 its gene organization and expression.";
 EMBO J. 5:2043-2049(1986).
 RL EMBO J. 5:2043-2049(1986).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 {RNA}(N).
 CC -1- SUBUNIT: IN CHLOROPLAST THE RNA POLYMERASE IS COMPOSED OF FOUR
 SUBUNITS: ALPHA, BETA, BETA', AND BETA".
 CC -1- SUBCELLULAR LOCATION: Chloroplast.

Query Match	6.3%	Score 97.5:	DB 1:	Length 680:
Best Local Similarity	23.7%	Pred. No. 2.8:		
Matches 49:	Conservative 25:	Mismatches 52:	Indels 81:	Gaps 13

RESULT	11
GARP_PLAUF	
ID	GARP_PLAUF
AC	P13816;
DT	01-JAN-1990 (Rel. 13, Created)
DT	01-JAN-1990 (Rel. 13, Last sequence update)
DT	15-JUN-1999 (Rel. 38, Last annotation update)
DE	Glutamic acid-rich protein precursor.

Plasmodium falciparum (isolate FC27 / Papua New Guinea).
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5837;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=69040048; PubMed=2903445;
RA Triglia T., Stal H.-D., Cretther P.E., Silva A., Anders R.F.,
Kemp D.J.
RA "Structure of a Plasmidom falciparum gene that encodes a glutamic
RT acid-rich protein (GAPP)".
RL Mol. Biochem. Parasitol. 31:199-202(1988).
RC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/sib-ch>).
CC or send an email to license@sib-sib.ch.

DR	EMBL; J03998; AAA29605.1; -.
DR	PIR; A54514; A54514.
KW	Repeat; Malaria; Antigen; Signal.
FT	SIGNAL 1 25

Query Match	6.28	Score 95.5	DB 1	Length 678
Best Local Similarity	20.88	Pred. No. 4		
Matches 55	Conservative 42	Mismatches 92	Indels 75	Gaps 10

Qy	43	NDNENIKKIOSOVNIKEAFISAAAEFTFSWYKSKDGEENELDKELKEGKIPAPLRSM	1020
Db	113	NENSVDKKDKKREKKHKKDKKEKE-----KKDKREKKDKREKKHKK-----	1544
Qy	103	FYTGCDYRDPLFGTDISKGHEGSKLKE--QIDSLFENQOQSPNGKTROEMWTEHSHET	1606
Db	155	-----EKHKKDKKKKENSEVMSLYKTGQHKPRNA-----TEHEEN	1911
Qy	161	WEAMICALVKGAKKD-----DEFTENYG-----YNNVFKSDKSTYLE--EFA	2000
Db	192	IDDEWVSFINNNAOGLLSSPYTRROGGCGIITSVHETSNPTKDNKDKNISEDKKEDH	2511
Qy	201	KRPQFLRWL-----TEWYDDYCYTRQYTKDKVOECKSNSDQJLCKDTECNKCEDYV	2511
Db	252	QOEEMLTLLDKEERKQKEKEKKEBEKITEKKKKKOEKEKKKOEKERRKQKEKERRKQKE--	3099
Qy	252	KYMKKKKKEWIPQDKYVYDDEPRKKR	275
Db	310	KEMKKQKK-TEKERRKKKEKKKKK	332

RESULT 12	
PKSL_BACSU	
ID	STANDARD;
PKSL_BACSU	PRT; 4427 AA

DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative polyketide synthase PKS1 (PKS1).
GN PKS1 OR PKSX OR PKSA OR OUTG.
OS *Bacillus subtilis*.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus
NCBI_TaxId=1423;

RP SEQUENCE FROM N.A.
RC STRAIN=168 / PB1424;
RY MEDIANE-03345834. PubMed-03345830.

RA Scotti C., Piatti M., Cuzzoni A., Perani P., Tognoni A., Grandi G.,
RA Galizzi A., Albertini A.M.;
RT "A *Bacillus subtilis* large ORF coding for a polypeptide highly
RT similar to polyketide synthases.",
RL Gene 130:65-71(1993).

RP	SEQUENCE FROM N.A
RC	STRAIN=168;

RA Kunst F., Ogatawara N., Moser T., Albertini A.M., Alloni G.,
RA Azevedo V., Belterio M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borris R., Boursier L., Brans A., Braun M., Brinell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Comerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dustsholt A., Ehlig S.D., Emmerson P.T.,
RA Etlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallizzi A., Galleon N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grand G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaetr-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Konhingsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lamber U., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Moore D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Portolillo S., Prescott A.M.,
RA Pirescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Senkan E., Schleich S., Schroeter R., Scottone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serron P., Shih B.S., Soldo B.,
RA Sotnick M., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi A., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Vial A., Wambuit R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zanchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256(1997).
RN [3]
RP SEQUENCE OF 3619-4427 FROM N.A.
RR STRAIN=168 / PBI424;
CC Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR
CC THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN
CC SECONDARY METABOLISM
CC -1- COFACTOR: CONTAINS 5 COVALENTLY BOUND PHOSPHOPANTETHEINES
CC (POTENTIAL).
CC -1- SIMILARITY: CONTAINS 5 ACYL CARRIER DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: 214098; CAA78479.1; -
DR EMBL: 011039; AAB85145.1; -
DR EMBL: 299113; CAA813602.1; -
DR EMBL: 235133; CAA84504.1; -
DR PIR: S25021; S25021.
DR Subtilist; B610698; pksL.
DR InterPro: IPR002198; ADH_short.
DR InterPro: IPR000794; Ketoacyl-synt.
DR InterPro: IPR003880; ppantte_attach.
DR Pfam: PF00106; adh_short; 1.
DR Pfam: PF00109; ketoacyl-synt; 4.
DR Pfam: PF00550; pp-binding; 5.
DR Pfam: PF02801; ketoacyl-synt; 4.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 5.
DR PROSITE: PS00606; B_KETOACYL_SYNTHASE; 1.
DR PROSITE: PS0075; ACP_DOMAIN; 5.
KW Transferrase; Acyltransferase; Antibiotic biosynthesis; NADP;
KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 211 280
FT DOMAIN 382 759
FT DOMAIN 937 1115
FT DOMAIN 1409 1602
FT DOMAIN 1687 1759
FT DOMAIN 1876 2253
FT DOMAIN 2491 2560
FT DOMAIN 2632 2701
FT DOMAIN 2823 3182
FT DOMAIN 3575 3776
FT DOMAIN 3854 3923
FT DOMAIN 4019 4373
FT BINDING 243 243
FT BINDING 1723 1723
FT BINDING 2523 2523
FT BINDING 2664 2664
FT BINDING 3886 3886
SQ SEQUENCE 4427 AA; 493398 MW; 9612521E561AB9F2 CRC64;

Query Match 6.2%; Score 95; DB 1; Length 4427;
Best Local Similarity 24.0%; Pred. No. 37;
Matches 43; Conservative 41; Mismatches 69; Indels 26; Gaps 8;
OY 110 RDLFTGDISKSGEESKLEQDISLFRKGDGKSPGKTRGEMWTEHSEITEALCALV 169
DB 1648 RAFLHET-----GIGKQGSINIKKSSLNQAEKADMGIDEITLNEKENVKQVLSVT 1703
OY 170 KIGAKKDDP---TENTGYNNVAFESDKSTLEE---FAK---RPOFLWLTWYDYCY 218
DB 1704 KIPAGGIDAEAPLEDYGDISIMIMHTYQLEKVFSGSLKTEFFEDIRSLRYRID--- 1760
OY 219 TRQYKLVQVQ--EKKSNDQKCDPECKKCDYKMKKKKEWIP--ODKYKDERDK 274
DB 1761 SRREKLIDLIGETGKPSVERKSEPE---KQELPIPRKSGFLPDQKEQVREKE 1814
RESULT 13
TB1 NOTCO
AC TB1 NOTCO STANDARD; PRT; 446 AA.
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tubulin beta-1 chain.
GN TUBB1.
OS Notothenia coriiceps neglecta (Black rockcod) (Yellowbelly rockcod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorphi; Acanthopterygii; Perciformes; Perciformes;
OC Notothenioidae; Nototheniidae; Notothenia.
OX NCBI_Taxid=8209;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-40.
RC TISSUE=Brain;
RX MEDLINE=93223253; PubMed=8467523;
RA Detrich H.W. III, Parker S.K.;
RT "Divergent neural beta tubulin from the Antarctic fish *Notothenia*
RT *coriiceps neglecta*: potential sequence contributions to cold
RT adaptation of microtubule assembly";
RL Cell Motil. Cytoskeleton 24:156-166(1993).
CC -1- FUNCTION: TUBULIN IS THE MAJOR CONSTITUENT OF MICROTUBULES. IT
CC BINDS TWO MOLES OF GTP, ONE AT AN EXCHANGEABLE SITE ON THE BETA
CC CHAIN AND ONE AT A NONEXCHANGEABLE SITE ON THE ALPHA-CHAIN.
CC -1- SUBUNIT: DIMER OF ALPHA AND BETA CHAINS.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO THE TUBULIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
DR EMBL: L08013; AAA49393.1; -
DR EMBL: S57698; AAB26110.1; -
DR PIR: A48407; A48407.
DR InterPro: IPR002453; Beta_tubulin.
DR InterPro: IPR000217; Tubulin.
DR InterPro: IPR003008; Tubulin_ftsz.
DR Pfam: PF00091; tubulin; 1.
DR PRINTS: PR01161; TUBULIN.
DR PROSITE: PS00227; TUBULIN_1;
DR PROSITE: PS00228; TUBULIN_B_AUTOREG; 1.
KW Microtubules; GTP-binding.
FT NP_BIND 140 146
SQ SEQUENCE 446 AA; 49806 MW; B8CE37298707D524 CRC64;
Query Match 6.1%; Score 94.5; DB 1; Length 446;
Best Local Similarity 25.8%; Pred. No. 2.9;

Matches 31; Conservative 18; Mismatches 46; Indels 25; Gaps 5;

QY 6 KNSNGYPMOCNINIVEDPRVC-MPPRROKLCVHFLANDNEIKLOSQVNIKEAFIKSA 64
 DB 336 KNSSEYEVMIJPNVKTAT---VCDIPRGLKMAFTIGNSTAIQELFKRIS--EQPTAMP 389
 QY 65 AAEFFSWYKSKDEG-----NELDKELKGGKTPPAFLSMFTFGDYRD 111
 DB 390 RRRKAFILHWY-----TGEQDEMEFTAESNMNDLVSEYOYODATAEERGEFEDEBYED 444

RESULT 14
 MN4_YEAST
 ID MN4_YEAST STANDARD; PRT: 1178 AA.
 AC P36044; P36043; P89095;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE MN4 protein.
 GN MN4 OR YKL200C/YKL201C.
 RX Saccharomyces cerevisiae (Baker's yeast).
 RA Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 RN Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RX MEDLINE=97175967; PubMed=9023541;
 RA Odani T., Shima Y.-I., Tanaka A., Jigami Y.;
 RT "Cloning and analysis of the MN4 gene required for phosphorylation
 of N-linked oligosaccharides in Saccharomyces cerevisiae."; Glycobiology 6:805-810(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Maia e Silva A., Bossier P., Villela C., Fernandes L., Soares H.,
 RA Guerrelto P., Rodrigues-Pousada C.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBD databases.
 CC -1- FUNCTION: MAY FUNCTION AS A POSITIVE REGULATOR FOR
 CC MANNOSE-6-PHOSPHATE TRANSFERASE. IS REQUIRED TO MEDIATE
 CC MANNOSE-6-PHOSPHATE TRANSFER IN BOTH THE CORE AND OUTER CHAIN
 CC PORTIONS OF N-LINKED OLIGOSACCHARIDES.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein (Probable).
 CC -1- SIMILARITY: TO YEAST YKR061W.
 CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
 CC FRAMESHIFTS, ONE OF WHICH PRODUCES TWO SEPARATE ORFS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to jlicens@sib-sib.ch).

DR EMBL: D83006; BA11676.1;
 DR EMBL: Z28201; CA82046.1;
 DR EMBL: Z28200; CA82044.1;
 DR PIR: S38037; S38037.
 DR PIR: S38038; S38038.
 DR SGD: S0001683; MN4.
 DR Transmembrane; Signal-anchor; Repeat.
 KM TRANSMEM 1 27
 FT DOMAIN 1 27
 FT TRANSMEM 28 48
 FT SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT LUMENAL (POTENTIAL).
 FT ARG/GLU/LYS-RICH (HIGHLY CHARGED).
 FT 17 X 8 AA TANDEM REPEATS OF K-K-K-K-E-E-
 FT E-E.
 FT REPEAT 1042 1049
 FT REPEAT 1050 1057
 FT REPEAT 1058 1065
 FT REPEAT 1066 1073
 FT REPEAT 1074 1081

FT REPEAT 1082 1089 6. (APPROXIMATE).
 FT REPEAT 1090 1097 7. (APPROXIMATE).
 FT REPEAT 1098 1105 8. (APPROXIMATE).
 FT REPEAT 1106 1113 9. (APPROXIMATE).
 FT REPEAT 1114 1121 10. (APPROXIMATE).
 FT REPEAT 1122 1129 11. (APPROXIMATE).
 FT REPEAT 1130 1137 12. (APPROXIMATE).
 FT REPEAT 1138 1144 13. (APPROXIMATE).
 FT REPEAT 1145 1152 14. (APPROXIMATE).
 FT REPEAT 1153 1160 15. (APPROXIMATE).
 FT REPEAT 1161 1168 16. (APPROXIMATE).
 FT REPEAT 1169 1174 17. (INCOMPLETE).
 FT REPEAT 37 40 POLY-TLE.
 FT DOMAIN 1042 1045 POLY-LYS.
 FT DOMAIN 1046 1049 POLY-LYS.
 FT DOMAIN 1050 1053 POLY-LYS.
 FT DOMAIN 1054 1057 POLY-LYS.
 FT DOMAIN 1058 1061 POLY-LYS.
 FT DOMAIN 1062 1065 POLY-LYS.
 FT DOMAIN 1066 1069 POLY-LYS.
 FT DOMAIN 1070 1073 POLY-LYS.
 FT DOMAIN 1074 1077 POLY-LYS.
 FT DOMAIN 1078 1081 POLY-LYS.
 FT DOMAIN 1082 1085 POLY-LYS.
 FT DOMAIN 1086 1089 POLY-LYS.
 FT DOMAIN 1094 1097 POLY-LYS.
 FT DOMAIN 1098 1101 POLY-LYS.
 FT DOMAIN 1102 1105 POLY-LYS.
 FT DOMAIN 1134 1137 POLY-LYS.
 FT DOMAIN 1157 1160 POLY-LYS.
 FT DOMAIN 1165 1168 POLY-LYS.
 SQ SEQUENCE 1178 AA; 139380 MW; BC05DAEDAFCB282 CRC64;

Query Match 6.1%; Score 94.5; DB 1; Length 1178;
 Best Local Similarity 22.1%; Pred. No. 9;
 Matches 66; Conservative 35; Mismatches 79; Indels 119; Gaps 12;

QY 32 RROKLCVHFLANDN--EIKLOSQVNIKEAFIKSAAEFFSWYK-----SKDEGN 83
 DB 858 RROKLEIYQCNLTFFIEKKRLHQLRINAPKLISSPAKDPFLGVEKAMWKDLSKSNQ 917
 QY 84 ELDEKELK-----EKKTPPAFLSMFTFGDYRDLEFGTDSKSGHGBGLKQIDSLK 137
 DB 918 TLDQVTKIHEEYVGR-----IDLE-----SLKYNPSLFN 950
 QY 138 NGDQSPNGKTRQEWMTESHEIWEAMLCALVIGAKKDDFTENYG-----YN 185
 DB 951 -----ITFDETGTITDNTEDYTPANTVEVNPVDFKS 982
 QY 186 NKFSDSKS-----TLEE-----FAKRPQFLRWLTWYDDCYT----- 219
 DB 983 NLNFSNSFLDLSYGLDFAPTLSDVNRKGIQMPDKPIIV-----YEDVAYAKLLEE 1036
 QY 220 ---RQYTKLVQEKCSNDQKCDTECNKCEYVYMKKKKEMIPQDKYKBERDKR 275
 DB 1037 RKRREKKKKKEEKKKKKEEKKKKKEEKKKKKEEKKKKKEE-----EKKKKKEEKKK 1092

RESULT 15
 SYL_TETTH
 ID SYL_TETTH STANDARD; PRT: 1081 AA.
 AC P36422;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 01-JUN-1994 (Rel. 29, Last annotation update)
 DE Isolation of tRNA synthetase (EC 6.1.1.5) (Isolation of tRNA ligase)
 DE (ILERS).
 GN ILA OR CUPE.
 OS Tetrahymena thermophila.
 OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
 OC Tetrahymenales; Tetrahymena.
 OX NCBI_TaxID=5911;
 RN [1]

RP SEQUENCE FROM N.A.
RX MEDLINE=92165815; PubMed=1371507;
RA Csank C., Martindale D.W.;
RT "Isolucyl-tRNA synthetase from the ciliated protozoan Tetrahymena
thermophila. DNA sequence, gene regulation, and leucine zipper
motifs."
RL J. Biol. Chem. 267:4592-4599(1992).
CC CATALYTIC ACTIVITY: ATP + L-Isoleucine + tRNA(Ile) = AMP +
CC diphosphate + L-isolucyl-tRNA(Ile).
CC -1 SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M30942; AAA30122.1; .
CC PIR: A42399; A42399.
CC HSSP: P56690; 11LE.
DR InterPro: IPR002300; tRNA-synt_1a.
DR InterPro: IPR001412; tRNA-synt_1.
DR InterPro: IPR002301; tRNA-synt_1le.
DR Pfam: PF00133; tRNA-synt_1; 1.
DR PRINTS: PR00984; TRNASYNT1LE.
DR TIGRFAMs: TIGR00392; 1le; 1.
DR PROSITE: PS00178; AA.TRNA.LIGASE.L; 1.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
FT SITE 53 "HIGH" REGION.
FT SITE 607 611 "KMSKS" REGION.
FT BINDING 610 610 ATP (BY SIMILARITY).
SQ SEQUENCE 1081 AA; 124850 MW; D4D72616AFB8C795 CRC64;

Query Match 6.1%; Score 94; DB 1; Length 1081;
Best Local Similarity 19.2%; Pred. No. 8.9; Mismatches 103; Indels 140; Gaps 15;
Matches 67; Conservative 39;

QY 37 CVHFLANDNEIKKIOSVNLKEA-----FIKSAAE-----TFFSW----- 72
DB 190 CATVLISN-----FETQONRYKEVDPSLFIATKTAEDPKTKFIATWTTPWTLPSNLAIV 243
QY 73 -----YYKSKDG-----EGNELDKELKGGKIPPAF 98
DB 244 NKQFDYVKYLDAKTOEHYLLAEGRLEPELYKKDKDGYKILEKFGSELYGREYEPLEPYFL 303
99 LKSMFTTF-----GDYRDFLEGTDI---SKHGSGSKLKEQIDSLFKNGDQKSP---NG-- 146
304 SRKODGCFRILAGDFTADAGTGVHCAPGFDDYKYSVANNITIKPDDPPVPVDENGAF 363
QY 147 -----KTRQE-----WTEHSHEIWEAMCALVK 170
DB 364 TINVSDPAGYITKEADKILRNKRLKRGILLVDSFKNYRPFQWRSDTPLYKAVHCWFIK 423
QY 171 IGAKKDDFTENTGYNNVKSSTLEEFARKPQPLRMILEWTDYCYTRQKY----- 223
DB 424 VFLAKDDLLAN-----NKKAYWVPKTAQEGREFNNWLN--VSDWCFSSRFWGNPIPI 474
QY 224 --LKDVQF-----KCKSNQIAKQDTECNKKCEQYVKKMKKKKEWIPQDK 265
DB 475 WSEDEFEVVCIGSVLELKKLTGATEITDLAKDFIDHILT-----IPSOK 518

Search completed: April 28, 2003, 10:30:26
Job time : 9.65193 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 10.3425 Seconds
(without alignments)
5498.552 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPRKNSNGYPPDMQGNIN.....KKWIPDKYKDERDKRRF 276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp Unclassified:*
15: sp_rv1rns:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1543	100.0	3542	5	Q9U5M2
2	750	48.6	1615	5	Q9U5M2
3	745	48.3	3287	5	Q8T325
4	634.5	41.1	2706	5	Q15870
5	614	40.4	3078	5	Q26031
6	614	39.8	182	5	Q00832
7	609	39.5	3026	5	Q26030
8	557.5	36.1	2658	5	Q8T5G0
9	544	35.3	169	5	Q25990
10	524	34.0	1685	5	Q9U4A2
11	516.5	33.5	2710	5	Q9X2B8
12	493.5	32.0	921	5	Q25989
13	493	32.0	2527	5	Q95W83
14	486	31.5	3006	5	Q26032
15	475	30.8	1711	5	Q96108
16	453.5	29.4	195	5	Q00831

17	447	29.0	176	5	Q9BUF0	Q9bjf0	plasmidium
18	439	28.5	184	5	Q9BUE6	Q9bj6	plasmidium
19	410.5	26.6	2042	5	Q25766	Q25766	plasmidium
20	408.5	26.5	173	5	Q25988	Q2588	plasmidium
21	408.5	26.5	2209	5	Q9U0G6	Q9u096	plasmidium
22	402.5	26.1	2135	5	Q61077	Q61077	plasmidium
23	396	25.7	2212	5	Q94657	Q94657	plasmidium
24	390.5	25.3	2228	5	Q60991	Q60991	plasmidium
25	385.5	25.0	2664	5	Q26033	Q26033	plasmidium
26	383	24.8	2209	5	Q97324	Q97324	plasmidium
27	382.5	24.8	2182	5	Q26034	Q26034	plasmidium
28	371.5	24.1	2647	5	P90580	P90580	plasmidium
29	369.5	23.9	2924	5	Q25733	Q25733	plasmidium
30	363.5	23.6	2163	5	Q9NFB6	Q9nfb6	plasmidium
31	353	22.9	5	Q9U0G5	Q9u095	plasmidium	
32	350	22.7	162	5	Q9BUE7	Q9bue7	plasmidium
33	347	22.5	160	5	Q9BUE9	Q9bue9	plasmidium
34	343.5	22.3	177	5	Q9BUE8	Q9bue8	plasmidium
35	341	22.1	2169	5	Q97312	Q97312	plasmidium
36	339.5	22.0	5	Q96296	Q96296	plasmidium	
37	299.5	19.4	1729	5	Q25734	Q25734	plasmidium
38	289.5	18.8	616	5	Q8T9N3	Q8t9n3	plasmidium
39	289.5	18.8	616	5	Q8T9L8	Q8t9l8	plasmidium
40	289.5	18.8	1210	5	Q96724	Q96724	plasmidium
41	289.5	18.8	1210	5	Q8W531	Q8w531	plasmidium
42	274	17.8	616	5	Q27361	Q27361	plasmidium
43	274	17.8	1475	5	Q25842	Q25842	plasmidium
44	270	17.5	616	5	P90585	P90585	plasmidium
45	269	17.4	616	5	Q25739	Q25739	plasmidium

ALIGNMENTS

RESULT 1

Q9U5M2 ID Q9U5M2 PRELIMINARY; PRT; 3542 AA.
AC Q9U5M2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2002 (TREMBLrel. 20, Last annotation update)
DE FCR3 CSA ligand (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=20006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheidig C., Baruch B., Oishi S., Fujii N.,
RA Fusai T., Parzy D., Miller L.H., Gysin J., Schert A.;
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
sulfate A: A receptor for human placental infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).
DR EMBL; AJ13811; CAB59840.1; -
DR InterPro: IPR001219; Neurotoxin.
DR Pfam: PF03011; PFEWP.
DR PRINTS; PR00284; TOXIN.
FT NON_TER 3542 3542
SQ SEQUENCE 3542 AA; 413089 MW; 970D85EE88DA2EC2 CRC64;

Query Match 100.0%; Score 1543; DB 5; Length 3542;
Best Local Similarity 100.0%; Pred. No. 6.7e-109;
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DCHPRKNSNGYPPDMQGNINLVEDPRVCMPRRQKLCVFLANDNEIKLQSOVNLKEAF 60
DB 1279 DCHPRKNSNGYPPDMQGNINLVEDPRVCMPRRQKLCVFLANDNEIKLQSOVNLKEAF 1338
QY 61 IKKAAATFFSWYKSKDDEGNELDKELKGIKIPAFLSMRYTFEDYDFLFGTDISK 120
|||||

Db 1339 IKSAAETFFSWYKSKDGEENELDEKELKGIIPAFILSMETFGDYRDLFGTDISK 1398
 OY 121 GHGEGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCALVIGAKKDDFTE 180
 Db 1399 GHGEGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCALVIGAKKDDFTE 1458
 OY 181 NYGYNNAKFSQSKSTTLLEEFARPOFLWLTWYDDCYTRQKYLKDVOEKCSNDOLKCD 240
 Db 1459 NYGYNNAKFSQSKSTTLLEEFARPOFLWLTWYDDCYTRQKYLKDVOEKCSNDOLKCD 1518
 OY 241 TECNKKCEDYVYKMKKKKEWIPDQKYYKDDERDKRRF 276
 Db 1519 TECNKKCEDYVYKMKKKKEWIPDQKYYKDDERDKRRF 1554

RESULT 2

O8T325 PRELIMINARY; PRT; 1615 AA.
 AC O8T325;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE PfEMP1 (Fragment).
 GN TM284VAR3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM284;
 RX MEDLINE=21927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 implicated in malaria in pregnancy."
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL; AJ420412; CAD20868.1; -
 FT NON_TER 1615 1615
 SQ SEQUENCE 1615 AA; 188639 MW; FCA896C008B6DEA6 CRC64;

Query Match

Best Local Similarity 48.6%; Score 750; DB 5; Length 1615;
 Matches 150; Conservative 33; Mismatches 79; Indels 32; Gaps 9;

OY 1 DCHPRKNSNGYPMOC---GNINLEVEDPRVCMPPRRQKLCVHFLANDNEIKLQSOVNL 56
 Db 1262 ECNVATTEGE-YPEWCECKKIQNIH---KGACMPRRRRLCVHFLKE---LKVEIYEKL 1312
 OY 57 KEAFIKSAAETFFSWYKSKDGEENELDEKELKGIIPAFILSMETFGDYRDLFGT 116
 Db 1313 KEAFIKSAAETFFSWYKSKDGEENELDEKELKGIIPAFILSMETFGDYRDLFGT 1372
 OY 117 DISKGHGEGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCAL----- 168
 Db 1373 DIGSDVG---NWNOKINAKLKIGSDSD--KRNMMNGIKEDVWKGWCGLSHDVSGNH 1427
 OY 169 ---VATGAKKDDFENNGYNNVKEFSK--STTLEEFARPOFLWLTWYDDCYTRQY 223
 Db 1438 KEIYAKKIMEDPQNNKYQNSVKTDEPSGTRKLEFAKVPFLWLTWYDDCYTRQY 1487
 OY 224 LKDVOEKCSNDOLKCDTECNKKCEDYVYKMKKKKEWIPDQKYYKDDERDKRRF 276
 Db 1488 LKEVESTCKSDOLKCDTECNKKCEDYVYKMKKKKEWIPDQKYYKDDERDKRRF 1541

RESULT 3

O8T326 PRELIMINARY; PRT; 3287 AA.
 AC O8T326;
 DT 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE PfEMP1 (Fragment).
 GN TM180VAR2.

OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TM180;
 RX MEDLINE=21927235; PubMed=11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 implicated in malaria in pregnancy."
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL; AJ420411; CAD20867.1; -
 FT NON_TER 3287 3287
 SQ SEQUENCE 3287 AA; 383550 MW; 58F8F866FC244536 CRC64;

Query Match

Best Local Similarity 48.3%; Score 745; DB 5; Length 3287;
 Matches 149; Conservative 34; Mismatches 79; Indels 32; Gaps 9;

OY 1 DCHPRKNSNGYPMOC---GNINLEVEDPRVCMPPRRQKLCVHFLANDNEIKLQSOVNL 56
 Db 1262 ECNVATTEGE-YPEWCECKKIQNIH---KGACMPRRRRLCVHFLKE---LKVEIYEKL 1312
 OY 57 KEAFIKSAAETFFSWYKSKDGEENELDEKELKGIIPAFILSMETFGDYRDLFGT 116
 Db 1313 KEAFIKSAAETFFSWYKSKDGEENELDEKELKGIIPAFILSMETFGDYRDLFGT 1372
 OY 117 DISKGHGEGSKLRQIDSLFKNGDQKSPNGKTRQEWMTESHEIWEAMLCAL----- 168
 Db 1373 DIGSDVG---NWNOKINAKLKIGSDSD--KRNMMNGIKEDVWKGWCGLSHDVSGNH 1427
 OY 169 ---VATGAKKDDFENNGYNNVKEFSK--STTLEEFARPOFLWLTWYDDCYTRQY 223
 Db 1438 KEIYAKKIMEDPQNNKYQNSVKTDEPSGTRKLEFAKVPFLWLTWYDDCYTRQY 1487
 OY 224 LKDVOEKCSNDOLKCDTECNKKCEDYVYKMKKKKEWIPDQKYYKDDERDKRRF 276
 Db 1488 LKEVESTCKSDOLKCDTECNKKCEDYVYKMKKKKEWIPDQKYYKDDERDKRRF 1541

RESULT 4

O15870 PRELIMINARY; PRT; 2706 AA.
 AC O15870;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DE PfEMP1 (Fragment).
 GN R29R4VAR1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IT 4/25/5;
 RX MEDLINE=97373957; PubMed=9230440;
 RA Rowe J.A., Moulds J.M., Newbold C.I., Miller L.H.;
 RT "Plasmodium falciparum rosetting is mediated by PfEMP1 and requires
 complement receptor 1."
 RL Nature 388:292-295(1997).
 DR EMBL; Y13403; CAA73831.1; -
 DR EMBL; Y13403; CAA73831.1; JOINED.
 DR InterPro; IPR002086; Aldehyde_dehydr.
 DR InterPro; IPR004258; PfEMP.
 DR PROSITE; PS00070; ALDEHYDE_DEHYDR_CYS; UNKNOWN_1.
 FT NON_TER 2706 2706
 SQ SEQUENCE 2706 AA; 308162 MW; 1C33D55AD5317D68 CRC64;

Query Match

Best Local Similarity 41.1%; Score 634.5; DB 5; Length 2706;
 Matches 131; Conservative 37; Mismatches 91; Indels 29; Gaps 10;

QY 1 DCHPRKNSNGYDPMOGNNINIVE-DRVCGMPRRKOLCVHFLANDNEIKKLOSOVNKEA 59
 DB 787 NCNARRKKN--EMQDDKMTFVGNNEGVCMPRRKKSICIHNLLEQOTK---NKYOLREA 841
 QY 60 FIKSAAEFFSWYKSKDGEGNEIDKEKIPPAFLRSMFYEGYDRDLFGDIS 119
 DB 842 FIKCAKKNLMDKKNKNEAEL---LKKKIPEDPRIRMEFYFGDFRDLCLNDMG 898
 QY 120 KGHGEGSKLEQIDSLFKNGDOKSPNGKTRQEWTEHSHEIWEAMCALVKGAK---- 174
 DB 899 K---DVDTKYNKINIKYFNNSKRGFKKIDPENWMNNGQIWMGMCALHADTKSIRK 955
 QY 175 KODFTNCGYNNKESDK-----STLEEPARKPOLRMLTEYDYCYTRKYLADVOE 229
 DB 956 KD---NYREKYTLAKRDGSGMGLSEPAKKPKFLRFMEVYDDYCKEROKRYTEVAS 1011
 QY 230 KCKSND--OLKCDTECNKCEYVYKMKKK--RMIPODKRYKRDCK 274
 DB 1012 TCKSIDGGLKCDRCGNCKCDEYKTKMKKKKEEMNLQDKTYKDKREK 1059

RESULT 5
 026031

PRELIMINARY: PRT: 3078 AA.

AC 026031: PRELIMINARY: PRT: 3078 AA.
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Variant-specific surface protein.
 GN VAR-1.
 OS Plasmodium falciparum (Isolate Dd2).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57267;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DD2;
 RX MEDLINE=95330813; PubMed-7606788;
 RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A., Peterson D.S., Ravetch J.A., Wellens T.E.;
 RT "The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-
 RT infected erythrocytes."
 RL Cell 82:89-100(1995).
 DR EMBL: LA0608; AAA75396.1;
 DR InterPro: IPR004258; PfEMP.
 DR Pfam: PF03011; PfEMP. 2.
 SQ SEQUENCE 3078 AA; 349297 MW; C8037C2BC3CCD7C3 CRC64;

Query Match 40.4%; Score 624; DB 5; Length 3078;

Best Local Similarity 43.8%; Pred. No. 7.5e-39; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPRKNSNGYDPMOG--GNINIVEPRVCMPPRRKOLCVHFLANDNEIKKLOSOVNKEA 59
 DB 1365 ECHPKES---YPMDCRKNNDISHD-GACMPRRKOLCVHFLANDNEIKKLOSOVNKEA 1420
 QY 60 FIKSAAEFFSWYKSK-DGEGNEIDKEKIPPAFLRSMFYEGYDRDLFGDIS 118
 DB 1421 FIKTAAEFFLSQWYKSKNDSEAKILDR---GLIPSGFLRSMYTFGDRDICTMTDI 1476
 QY 119 SKHGEGSKLEQIDSLFKNGDOKSPNGKTRQEWTEHSHEIWEAMCALVKGAKKD- 177
 DB 1477 SKQONVAAKAKDIGFFSKDGSGSLSRQEMWKTNGEIKMGMCALTKVYTDITDK 1536
 QY 178 --FTENGYNNKESDK-STLEEPARKPOLRMLTEYDYCYTRKYLADVOEKCS- 233
 DB 1537 RKIKNDYSDKVNQNGNSLEEFPAKPOLRMLTEYDYCYTRKYLADVOEKCS- 1596
 QY 234 NDOLKCDT---ECNKCEDYVYKMK-KKKEMI-----PDQKYYK 268
 DB 1597 NSTQOCNDAKHRCNOACRAVOEYVENKKRFSQGTNMFVLKANVQPDPEYK 1648

RESULT 6

ID 000832 PRELIMINARY: PRT: 182 AA.

AC 000832:
 DT 01-JUL-1997 (TReMBLrel. 04, Created)
 DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Partial erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum (Isolate Dd2).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57267;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DD2;
 RX MEDLINE=97342604; PubMed-9199301;
 RA Fischer K., Horrocks P., Preuss M., Wiesner J., Wensch S., Camargo A.A., Lanzer M.;
 RT "Expression of var genes located within polymorphic subtelomeric domains of Plasmodium falciparum chromosomes."
 RL Mol. Cell. Biol. 17:3679-3686(1997).
 DR EMBL: Y11910; CAA72661.1;
 FT NON_TER 1
 FT NON_TER 1
 SQ SEQUENCE 182 AA; 20830 MW; 197C4A677BAD148B CRC64;

Query Match 39.8%; Score 614; DB 5; Length 182;

Best Local Similarity 63.7%; Pred. No. 1.7e-39; Mismatches 16; Conservative 28; Mismatches 34; Indels 4; Gaps 4;

QY 31 PROKLCVHFLANDNEIKKLOSOVNKEAFIKSAAEFFSWYKSKDGE-GNELDKEL 89
 DB 1 PROKLCVHFLANDNEIKKLOSOVNKEAFIKSAAEFFSWYKSKDGE-GNELDKEL 60
 QY 90 KEKIPPAFLRSMFYEGYDRDLFGDISKSGEGSKLEQIDSLF-KRGDDKSPNGKT 148
 DB 61 KKGITPEFLRSMFYEGYDRDLFGDISKSGEGSKLEQIDSLF-KRGDDKSPNGKT 120
 QY 149 ROEWTEHSHEIWEAMCALVKI-GAKKDDFTENGYNNKESDKS-TTLEEPARKPOL 206
 DB 121 REDWMTKNGPDIMWGMICALTNASGAKKETLTDKYYKESVTFGDSGPNLQTFSSRQFL 180
 QY 207 RW 208
 DB 181 RW 182

RESULT 7
 026030
 ID 026030 PRELIMINARY: PRT: 3026 AA.
 AC 026030:
 DT 01-NOV-1996 (TReMBLrel. 01, Created)
 DT 01-NOV-1996 (TReMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE Variant surface protein (fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE OF 2044-2922 FROM N.A.
 RC STRAIN-IT 4/25/5;
 RX MEDLINE=95330813; PubMed-7606788;
 RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A., Peterson D.S., Ravetch J.A., Wellens T.E.;
 RT "The large diverse gene family var encodes proteins involved in cytoadherence and antigenic variation of Plasmodium falciparum-
 RT infected erythrocytes."
 RL Cell 82:89-100(1995).
 DR EMBL: LA0608; AAA75396.1;
 DR InterPro: IPR004258; PfEMP.
 DR Pfam: PF03011; PfEMP. 2.
 SQ SEQUENCE OF 2044-2922 FROM N.A.
 RC STRAIN-IT 4/25/5;
 RA Smith J.D., Chlitis C.E., Craig A.G., Roberts D.J.,

RA Hudson-Taylor D.E., Peterson D.S., Pinches R., Newbold C.I.,
 RA Miller L.H.;
 RT "Switches in the expression of Plasmodium falciparum var genes
 RT correlate with changes in antigenic and cytoadherent phenotypes of
 RT infected erythrocytes.";
 RL Cell 0:0-0(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IT 4/25/5;
 RX MEDLINE=9904502;
 RA Smith J.D., Kyes S., Craig A.G., Pagan T., Hudson-Taylor D.,
 RA Miller L.H., Baruch A.I., Newbold C.I.;
 RT "Analysis of adhesive domains from the A4VAR Plasmodium falciparum
 RT erythrocyte membrane protein-1 identifies a CD36 binding domain";
 RL Mol. Biochem. Parasitol. 97:133-148(1998).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-IT 4/25/5;
 RA Smith J.;
 RT Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
 [5]
 SEQUENCE FROM N.A.
 RC STRAIN-IT 4/25/5;
 RA Kyes S., Smith J.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: I42244; AAD03351.1;
 DR InterPro: IPR004258; PfEMP.
 DR Pfam: PF03011; PfEMP; 2.
 FT NON_TER 3026
 SQ SEQUENCE 3026 AA; 348124 MW; EC0A971A82193788 CRC64;

Query Match 39.5%; Score 609; DB 5; Length 3026;
 Best Local Similarity 42.3%; Pred. No. 1e-37; Mismatches 89; Indels 34; Gaps 10;
 Matches 127; Conservative 50;

QY 1 DCHPK--KNSNGYPMOCGNNILVEDPRVCMPPRRQKLCVHFLANDNEIKLOSQVNLKE 58
 DB 2005 ECNTKATYPPKNDYPCWNCCTDKVYNREGSCMPRRQKLCIH--NLEHSEKATELELRK 2061
 QY 59 AFKSAATETFPFWYK-----SKDGE-----NEIDKELKSKTIPAPFLRSMTYTE 106
 DB 2062 AFTECAITETFMIDKYEKKDEKTEGGGISDPPDQKLEGTTPEDFKROMFYTY 2121
 QY 107 GYRDFLEFGTIDISKGEGSKLEQIDSLFKNGDQSPKGTROBMTWHEHSEIWEAMIC 166
 DB 2122 GYRDFLEFGTIDISKGEGSKLEQIDSLFKNGDQSPKGTPTWMDYCPDIWKNVC 2181
 QY 167 AL---VKIGAKKDDPTENTGYNNVFSKSTTLEEFANRPOFLRWLTWYDYCYTRQKY 223
 DB 2182 GLSHHIKNG--NEQOLRKNLTDNN--KYTKISSKLEDFASRPOFLRWFIEGDFCGERVYK 2239
 QY 224 LKDVQEKCKSNOLKQDTE-----CNKKCEDYVKKKKKKEMIPDP--KYKDERDKK 274
 DB 2240 INQLTGC---NEYECGSGOENCKKCAKKNACCAVYSKWLKMDQYEQOTARFKDKDKK 2296

RESULT 8

Q8T5G0 PRELIMINARY; PRT; 2658 AA.
 AC Q8T5G0;
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Erythrocyte membrane protein-1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-JUP8;
 RA Chattopadhyay R., Pillai C.R., Chitnis C.;
 RT "Identification of a domain responsible for binding to intercellular

RT adhesion molecule-1 from a Plasmodium falciparum field isolate";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY028643; AAK49742.1;
 FT NON_TER 2658
 SQ SEQUENCE 2658 AA; 301326 MW; ABE33BC9167E7B7 CRC64;

Query Match 36.1%; Score 557.5; DB 5; Length 2658;
 Best Local Similarity 42.7%; Pred. No. 7.5e-34;
 Matches 123; Conservative 43; Mismatches 85; Indels 37; Gaps 12;

QY 2 CHPK-KNSN-GYDDMOCGNNILVEDP-RVCMPPRRQKLCVHFL--ANDNEIKLOSQV 54
 DB 1326 CNAKININIESIPDWTCVNSKPEENNEYGCPMPRRQKLCVLYLKLGEDEQK----- 1379
 QY 55 NLKAFIKSAAEETFPFWYKSKDGEINELD--KEIKESKIPAPFLRSMTYFGYRDFL 113
 DB 1380 -FKDAFIKTAAEETFLSMQYKSK--NSMDIKLOSQSEIPEERLRSMYFYGYRDF 1434
 QY 114 FGDISKSGEGSKLEQIDSLFKNGDQSPKGTROBMTWHEHSEIWEAMICAL--VK 170
 DB 1435 LMTDISKKEGDVSDAKGKIDAVFNKRYTDTN-----RTKMDTNGPEIWEAMICALTHGVT 1489
 QY 171 IGAKKDDPTENTGYNNVFSKSTTLEEFANRPOFLRWLTWYDYCYTRQKYLKDVQEK 230
 DB 1490 NTDNKRKIKTIDVSYKELQSKNTTLEKFAERPOFLRWFIEGDFCGERQKKYNELKEK 1549
 QY 231 CK-----SMDQLKCD--TECNKKCEDYVKKKKKE--WIPDQKYK 268
 DB 1550 CNKCCNGNNGVTSDECKTCVBOCKCEYKGFITEMQNMNKKRYK 1597

RESULT 9

Q25990 PRELIMINARY; PRT; 169 AA.
 AC Q25990;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
 DE (clone pM4) ORF (Fragment).
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DD2/NM;
 RX MEDLINE=95350219; Pubmed=7624377;
 RA Peterson D.S., Miller L.H., Wellens T.E.;
 RT "Isolation of multiple sequences from the Plasmodium falciparum genome
 RT that encode conserved domains homologous to those in erythrocyte-
 RT binding proteins";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
 DR EMBL: L38455; AAC37241.1;
 FT NON_TER 1
 FT NON_TER 169
 SQ SEQUENCE 169 AA; 19120 MW; BCG6B0FC3EE1380 CRC64;

Query Match 35.3%; Score 544; DB 5; Length 169;
 Best Local Similarity 62.1%; Pred. No. 3.5e-34;
 Matches 105; Conservative 26; Mismatches 34; Indels 4; Gaps 4;

QY 38 VHFLLANDNEIKLOSQVNLKEATKSAAEETFPFWYKSKDGE--GNELDKELKGRIPP 96
 DB 1 VHFLLANDNEIKLOSQVNLKEATKSAAEETFPFWYKSKDGE--GNELDKELKGRIPP 60
 QY 97 AFLRSMTYFGYRDFLEFGTIDISKGEGSKLEQIDSLF--KNGDQSPKGTROBMTWHE 155
 DB 61 EFLRSMTYFGYRDFLEFGTIDISKGEGSKLEQIDSLF--KNGDQSPKGTROBMTWHE 120
 QY 156 HSEIWEAMICALVKI--GAKKDDPTENTGYNNVFSKDS--TLEEFANR 202
 DB 121 NGPDWIKMICALTNASGAKKETLTDKRYESVTEGDNQPNLQFSSR 169

```

RESULT 10
ID 0904A2 PRELIMINARY; PRT; 1685 AA.
AC 0904A2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2002 (TREMblrel. 20, Last annotation update)
DE Variant surface protein PfEMP1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IT;
RA MEDLINE=20144115; PubMed=10677532;
RA Smith J.D., Grais A.C., Krick N., Hudson-Taylor D., Kyes S., Fagen T.,
RA Pinches R., Baruch D.I., Newbold C.I., Miller L.H.;
RT Identification of a Plasmodium falciparum intercellular adhesion
RT molecule-1 binding domain: A parasite adhesion trait implicated in
RT cerebral malaria.
Proc. Natl. Acad. Sci. U.S.A. 97:1766-1771(2000).
DR EMBL: AF193424; AAF1980.1; -.
DR InterPro: IPR000454; Eub.ATPase_Csub.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 1.
DR PROSITE: PS00605; ATPase_C; 1.
FT NON_TER 1685 1685
SQ SEQUENCE 1685 AA; 193671 MW; D1FDA266680551E CRC64;

Query Match 34.0%; Score 524; DB 5; Length 1685;
Best Local Similarity 42.7%; Pred. No. 1.6e-31;
Matches 122; Conservative 35; Mismatches 89; Indels 40; Gaps 11;

QY 14 WOCGINIVVDPVYVCMPPRRQKLCVHFLANDNEIKKIQSQVNLKEAFIKSAAEFFESWY 73
DB 1252 WKCENTKLGDEGVCMPPRRQNCVHLYTKLNDISK--EEDLEAFIKSAAEFFLLRQ 1308
QY 74 YVSKDQEGNELDKELKGIIPPAFLRSMFTYFGDYRDFGTDISK--SHGESEKLEQ 131
DB 1309 YVSKNVED--DKILHRDMPPEFPFRSMFTYFGDYRDLIDLDISKIADHVTAKK 1365
QY 132 ISLFLKNGDOKSPNGK---TRQEWTEHSHEIWEAMICAL-VKIGAKRD---FTENY 182
DB 1366 ITAVFQKIGSKTNGKVTLEREGMKEYGSLIMGICALSYLNTETKMGEGVATYLMKY 1425
QY 183 GYNNVFSKSTLEEFARQPOFLRMTLEWYDYCYTRQKYLNDVQKCS-----ND 235
DB 1426 IYKN---NDIKYLEEPASPPPLRWTEGEFVKNRKKELVSLKRRCDSCITLRNGTS 1482
QY 236 QLKCDPE-----CNKCEDVYVKKMKKKKEWIPQDKYKDRDKRRF 276
DB 1483 NKTCDDNENGACKTQCEKTKKMERWK-----KHYSQ--KKKF 1520

RESULT 11
ID 09XZB8 PRELIMINARY; PRT; 2710 AA.
AC 09XZB8;
DT 01-NOV-1999 (TREMblrel. 12, Created)
DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Variant-specific surface protein.
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CS2;
RA MEDLINE=99238507; PubMed=10220443;
RA Reeder J.C., Cowman A.F., Davern K.M., Beeson J.G., Thompson J.K.,
RA Rogerson J.J., Brown G.V.;

```

```

RT "The adhesion of Plasmodium falciparum-infected erythrocytes to
RT chondroitin sulfate A is mediated by P. falciparum erythrocyte
RT membrane protein 1."
RL Proc. Natl. Acad. Sci. U.S.A. 96:5198-5202(1999).
DR EMBL: AF134154; AAD29126.1; -.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 2.
SQ SEQUENCE 2710 AA; 309247 MW; 0AC64F0A35BF512 CRC64;

Query Match 33.5%; Score 516.5; DB 5; Length 2710;
Best Local Similarity 40.4%; Pred. No. 1e-30;
Matches 118; Conservative 43; Mismatches 100; Indels 31; Gaps 12;

QY 2 CHPKNSNGYPPDMQCGNINIVDPVYVCMPPRRQKLCVHFLANDNEIKKIQSQVNLKEAFI 61
DB 918 CNPRE--YPMDCCKNNIDNSHGACMPRRQKLCVRLDLOGEIKRPE---DILKFI 971
QY 62 KSAAEFFESWYVYKSKDGESENELDKELKGIIPPAFLRSMFTYFGDYRDFLEGTDIS-- 119
DB 972 NCAKETHFAMHMYK-KDYNNAE--NELKSGKIDEGFRKOMYTFGDFRIFFGTDISSC 1028
QY 120 --KGHGESEKIKRQIDSLFKNGDOKSPNGKTRQEWTEHSHEIWEAMICALVYK--IGAK 174
DB 1029 RYIKDVSQITKSLGQDATTETKGTIDNNKKELEWTTIGPKIWEGLCALTNGLSESE 1088
QY 175 KDFETENYGYNNVFSDK-STLEEFARQPOFLRMTLEWYDYCYTRQKY---LKDY--- 227
DB 1089 KKNILDYSYNNKLNNAKDKDCCLEKFSKQPLFLWYEWSEDFREKREKLEDEKVEDCIR 1148
QY 228 ---QEKCSNDQKCDTECKKCEYVYK-KKEWIPD-KYKDERDK 274
DB 1149 AKDEGCKNN--KSNNSCVKCEYENYITGKTYQESQEGKENTEKROKK 1197

RESULT 12
ID 025989 PRELIMINARY; PRT; 921 AA.
AC 025989;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE (clone p531H) ORF (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DD2/NM;
RA MEDLINE=95350219; PubMed=7624377;
RA Peterson D.S., Miller L.H., Wellens T.E.;
RT "Isolation of multiple sequences from the Plasmodium falciparum genome
RT that encode conserved domains homologous to those in erythrocyte-
RT binding proteins."
Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
DR EMBL: L38454; AAC37240.1; -.
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 1.
FT NON_TER 1
SQ SEQUENCE 921 AA; 104711 MW; D98BC07763628746 CRC64;

Query Match 32.0%; Score 493.5; DB 5; Length 921;
Best Local Similarity 40.0%; Pred. No. 1.7e-29;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPKNSNGYPPDMQCGNINIVDPVYVCMPPRRQKLCVHFLANDNEIKKIQSQVNLKE 58
DB 414 EYRKET---YSEWTCSESKIKMGHGACIPPRQKLCVHL-----EKIMTNTNELKY 464
QY 59 AFKSAAEFFESWY-YVSKDQEGNELDKELKGIIPPAFLRSMFTYFGDYRDFLEGTD 117
DB 465 AFKICAAAEFFFLMONGNKKKNGAEDLDEKLGITIPEDFKRMFTFPADYRICGTD 524
QY 118 I-----SKGHGESEKLEQIDSLFKNGDOKSPNG-KTRQEWTEHSHEIWEAMICAL- 168

```

```

Db 525 ISSKDTSGVGV---KVCNIDVDF-----YKISNIRKSKWMEHNGVINEGMLCALSY 577
Qy 169 -----VKIGAKKDDFTENYGNANKF--SDKSTILEEAKRPOLRWLTMYDYCYTRO 221
Db 578 DTSLNANVPETIKLTTEGNNNNEKVIYFGSDSTLSTKSEBPOFLRWLTENGNECKCKOK 637
Qy 222 KYLKVOECKSKNDQKCDTECNK--CEDYVK-----YMKKKEMLPQDK- 265
Db 638 KEYKVLAKCKDCD-VDDGCKNGCKVACKQCKQYHSWIGIWDNKKKQGRTEVKKI 696
Qy 266 -YKDERDK 274
Db 697 PLKEDKDKV 706

```

RESULT 13
ID 095W83 PRELIMINARY; PRT: 2527 AA.

```

NC 01-DEC-2001 (Tremblrel. 19, Created)
DE 01-DEC-2001 (Tremblrel. 19, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
GN Erythrocyte membrane protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=21442075; Pubmed=11557894;
RA Flick K., Scholander C., Chen Q., Fernandez V., Pouvelle B., Gysin J.,
RA Wehleren M.;
RT "Role of nonimmune IgG bound to PfEMP1 in placental malaria.";
RL Science 293:2098-2100(2001).
DR EMBL: AF366567; AAL12845.1;
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 2.
FT NON_TER 2527
SQ SEQUENCE 2527 AA; 292137 MW; 59E2766BF0425E4 CRC64;

```

Query Match Best Local Similarity 32.0%; Score 493; DB 5; Length 2527; Matches 116; Conservative 50; Mismatches 95; Indels 76; Gaps 12;

```

Qy 1 DCHPKKNSG---YPDGCGNINL---VDPVCMPPRRQKLCVFLANDNEIKKLSQV 54
Db 1663 DACEEYKNGKTAAYPGMKCTTKIEKGEDDYVCIIPRRQKLYT-----KQKFEFGSK 1716
Db 55 N--LKEAFKSAATEFESWY---KSKDGEENDELKEL 89
Db 1717 NDELKRAFLTECAAVETFEFAMHKYKEDKNEKEKETSILVRLGFLGKQOSSPDERAOKDL 1776
Qy 90 KEGKIPPAFLRMETFGYRDLFGTIDISKHGEGR-LKEQIDSLFRKNDOKSPNGKT 148
Db 1777 EGGKIPDEKRRMFTYFGYRDLFGTIDISKHGEGR-LKEQIDSLFRKNDOKSPNGKT 148
Qy 149 ROEWMTSHETIWEAMLCAL-----VKIGAKKDDFTENYGNANKF--SDKSTILEEAK 201
Db 1831 RETWNNNAKDIIMNGMICALSYDTEKRIKIGIYQLTQNDYKVTIDNNTLEIYAT 1890
Qy 202 RPOFLRWLTMYDYCYTROKYLKDVQCKS-----NDOLKCD----- 240
Db 1891 VPQFLRSEEWAEFCRRKHKHLAIHKDCRSDPFGMKYCSGDGHDGNEELKINDMAD 1950
Qy 241 ---TECNKCEDYKVMKK-KEMIPQDKYKDERDK 273
Db 1951 SYCPDCKAKCKRYNWKIEKVEEYFNQSKYKKEHLK 1987

```

RESULT 14
ID 026032 PRELIMINARY; PRT: 3006 AA.

```

AC Q26032;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
DE Variant-specific surface protein.
GN VAR-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
RX SEQUENCE FROM N.A.
RX STRAIN=FCR3;
RX MEDLINE=95330813; Pubmed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellem T.E.;
RT "The large diverse gene family var encodes proteins involved in
RT cytoadherence and antigenic variation of Plasmodium falciparum-
RT infected erythrocytes."
RL Cell 82:89-100(1995).
DR EMBL: LA0609; AAK75397.1;
DR InterPro: IPR004258; PfEMP.
DR Pfam: PF03011; PfEMP; 2.
SQ SEQUENCE 3006 AA; 343769 MW; 8FD5FF475F943C74 CRC64;

```

Query Match Best Local Similarity 31.5%; Score 486; DB 5; Length 3006; Matches 115; Conservative 39; Mismatches 108; Indels 32; Gaps 13;

```

Qy 5 KNSNGYDPMOCNINLVEDPVCMPRRQKLCVFLANDNEIKKLSQVNLKEAFKISA 64
Db 1247 KAGKDYKGMPCNSQIHHTHNGACMPRRQKLCVSGLTDRITALE---YIRFEFKISA 1303
Qy 65 AAEFTFSWYKSKDGEENELDKLEKGIIPAPLRSMTYFGYRDLFGTIDISKHGE 124
Db 1304 ALETHFAMDRYKENGEA---EAEIKNGNIPEGFRKQYVYFGYRDLFGTIDISKHGE 1359
Qy 125 GSKLEQIDSLF-KNGDQK-----SPNGKTRQEWMTSHETIWEAMLCALY-KIG--AKK 175
Db 1360 ISGVSPVITILEKENAKYAKONSNNELDDWDGDKGIMGICALTHKISDEKK 1419
Qy 176 DDFTEYGNVANKSDK-STLLEFPARRPOLRWLTMYDYCYTROKYLKDVQCK- 231
Db 1420 KEINKKSYKLLNSPGRKNEVEDFAKPPFLRFIEMGDEFCQOREKRAKYKVSQSDA 1479
Qy 232 KSNOLKCDTECNK-----EDYKVMKKKKEMIPQDKYKDER--DKKRF 276
Db 1480 KDYDGC-NRKSNSVCSACKVYEDYI--TKKVEYTKQKGFDAEKIDKEGY 1530

```

```

RESULT 15
ID 096108 PRELIMINARY; PRT: 1711 AA.
AC 096108;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
DE 01-DEC-2001 (Tremblrel. 19, last annotation update)
GN PfEMP1.
GN PFB0010W.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=5833;
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE=99021743; Pubmed=9804551;
RA Gardner M.J., Tetteh H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shallow S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Pettes M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Frazer C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum."
RL Science 282:1126-1132(1998).
DR EMBL: AE001366; AAC71792.1;

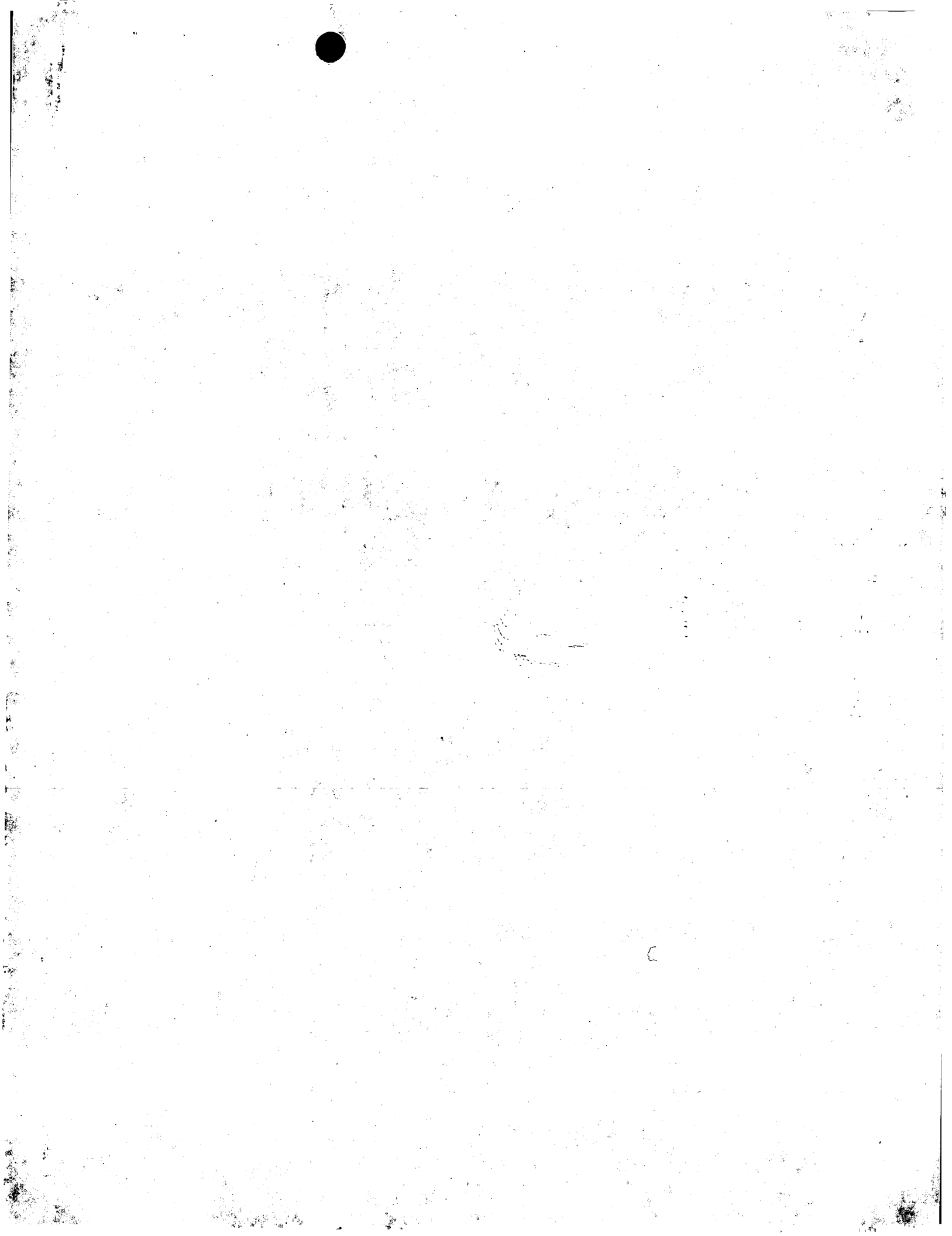
```

DR InterPro: IPR004258; PfEMP.
 DR Pfam: PF03011; PfEMP; 1.
 SQ SEQUENCE 1711 AA; 196550 MW; 7BBDE6C52742235B CRC64;

Query Match 30.8%; Score 475; DB 5; Length 1711;
 Best Local Similarity 38.0%; Pred. No. 9e-28;
 Matches 116; Conservative 39; Mismatches 86; Indels 64; Gaps 16;

QY 14 MOC---GNINLVDPVPCVPPRRQKLCVHFLANDNEIKKIQ--SOVNLKEAFIKSAAET 68
 Db 856 WRCIAPSGTSSGKDGALCVPRQELCIYLL-----KELSDPTQKGLRFAFKITAQET 909
 QY 69 FESWYIK-SKGE--GNEID-----KELKEGKIPPAFLKSMFYTFGDRDLFGDISK 120
 Db 910 YILMQYKKEKQNETASTELDIDDPQTOINGEIPEDFKRMFTYTFGDYRDLFLGRYIG- 968
 QY 121 GHGEGSKLEQIDSLFKNGDKSPNG---KTRQEWMTESHEIWEMLCALWKIGAKKD 176
 Db 969 --NDLKVNNNTAVFONGDH-IPNGOKTDROKQEFWGTGKDIMKGLCALQDAGGKR- 1024
 177 DPTENYGYNNVVKFSD--KSTTLEEFARPOFLMLTEWYDDYCYTRQKYLKDVQEC--- 231
 Db 1025 TLTFETYNYSVTFNGHLTGKLNELFASRPSFLRMWTEWMDQFCRERITQQLIKRCMYY 1084
 QY 232 ----KSNQDLKCDTECNKCCEDYKYMKKKKEMIP--QDKY-----YKD 269
 Db 1085 QYNGDKGKDDKK--EKCTEACTYY-----KMLTNMQDNYKKONQRYTEVKGTSPLYKE 1135
 QY 270 ERDCK 274
 Db 1136 DSDVK 1140

Search completed: April 28, 2003, 10:35:35
 Job time : 18.3425 secs



GenCore version 5.1.4_p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 2.78453 Seconds
(without alignments)
2916.372 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 1543
Sequence: 1 DHPKNSNGIPDMOCNIN.....KKWIPDKYDERDKRRE 276

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

tal number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :

1: /cgcn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgcn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgcn2_6/prodata/1/1aa/5A.COMB.pep:*
4: /cgcn2_6/prodata/1/1aa/5B.COMB.pep:*
5: /cgcn2_6/prodata/1/1aa/PCTUS.COMB.pep:*
6: /cgcn2_6/prodata/1/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	624	40.4	2710	2	US-08-568-459A-12
2	624	40.4	2710	2	US-08-487-826B-12
3	624	40.4	2710	4	US-09-210-288-12
4	624	40.4	3060	2	US-08-487-826B-14
5	493.5	32.0	921	2	US-08-568-459A-8
6	493.5	32.0	921	2	US-08-487-826B-8
7	493.5	32.0	921	4	US-09-210-288-8
8	382.5	24.8	2182	2	US-08-487-826B-16
9	274	17.8	1435	2	US-08-568-459A-4
10	274	17.8	1435	2	US-08-487-826B-4
11	274	17.8	1435	4	US-09-210-288-4
12	273.5	17.7	311	2	US-08-568-459A-21
13	273.5	17.7	311	4	US-08-487-826B-33
14	273.5	17.7	311	4	US-09-210-288-21
15	260	16.9	749	2	US-08-568-459A-6
16	260	16.9	749	2	US-08-487-826B-6
17	260	16.9	749	4	US-09-210-288-6
18	223	14.5	1115	2	US-08-568-459A-2
19	223	14.5	1115	2	US-08-487-826B-2
20	223	14.5	1115	6	US-09-210-288-2
21	223	14.5	1115	6	5198347-6
22	218.5	14.2	324	2	US-08-568-459A-17
23	218.5	14.2	324	2	US-08-487-826B-29
24	218.5	14.2	324	4	US-09-210-288-17
25	210	13.6	700	2	US-08-568-459A-10
26	210	13.6	700	2	US-08-487-826B-10
27	210	13.6	700	4	US-09-210-288-10

28	172.5	11.2	197	6	5198347-2	Patent No. 5198347
29	172.5	11.2	778	6	5198347-4	Patent No. 5198347
30	164	10.6	277	2	US-08-568-459A-15	Sequence 15, Appl
31	164	10.6	277	4	US-08-487-826B-27	Sequence 27, Appl
32	164	10.6	277	4	US-09-210-288-15	Sequence 15, Appl
33	142	9.2	282	2	US-08-568-459A-16	Sequence 16, Appl
34	142	9.2	282	2	US-08-487-826B-28	Sequence 28, Appl
35	142	9.2	282	4	US-09-210-288-16	Sequence 16, Appl
36	140	9.1	291	2	US-08-568-459A-13	Sequence 13, Appl
37	140	9.1	291	2	US-08-487-826B-25	Sequence 25, Appl
38	140	9.1	291	4	US-09-210-288-13	Sequence 13, Appl
39	133.5	8.7	411	2	US-08-568-459A-19	Sequence 19, Appl
40	133.5	8.7	411	2	US-08-487-826B-31	Sequence 31, Appl
41	133.5	8.7	411	4	US-09-210-288-19	Sequence 19, Appl
42	126	8.2	411	2	US-08-568-459A-20	Sequence 20, Appl
43	126	8.2	411	4	US-08-487-826B-32	Sequence 32, Appl
44	126	8.2	411	4	US-09-210-288-20	Sequence 20, Appl
45	123	8.0	362	2	US-08-568-459A-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-568-459A-12
Sequence 12, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OR INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-12
Query Match 40.4%; Score 624; DB 2; Length 2710;

Best Local Similarity 43.8%, Pred. No. 8.5e-53;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

```
QY 1 DCHPRKNSNGEYPMOC--GNINLVDPVCMPPRRKLCVHLANDNEIKLOSQVNLKEA 59
DB 1367 ECNPKES---YPMDCNKNIDISHD-GACMPRRKLCVYIAHSQENIKTDDNLKDA 1422
QY 60 FIKSAAEFFSWYKSK-DGEGNELDKELKGRIPAPLRSMEYTGDFYDFLFGDI 118
DB 1423 FIKTAAEFFLSWQYKSKNDSEAKILDR---GLIPQFLASMAVTFGDRDCLMTDI 1478
QY 119 SKHGSGSLKQIDSLFKNQDQSPNGKTRQEWMTESHSIHWAMCALYKIGAKKD- 177
DB 1479 SKQNDVAAKADKIGKFFSKDSKSPGSLRQEWMTKNGPEIWMGMLCALYKVTDTDNK 1538
QY 178 --FTENYGNVNFSDK-STLLEFAKRPQFLRWLTWYDDCYTRQKYLKDVQEKCS- 233
DB 1539 RKINDYSYDKVNOSONGPNLSLEFPAKQFLRWMIEMGEFCERKKNIIKDACNEI 1598
QY 234 NDOLKCDT---ECNKKCEDYKVKM-KRKEWI-----PODKYYK 268
DB 1599 NSTQOCNDAKRCNQACRAIOEYENKKEFSQTNNEVLKANYQPDPEYK 1650
```

RESULT 2

US-08-487-826B-12
Sequence 12, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487, 826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-12

Query Match

Best Local Similarity 40.4%; Score 624; DB 2; Length 2710;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

```
QY 1 DCHPRKNSNGEYPMOC--GNINLVDPVCMPPRRKLCVHLANDNEIKLOSQVNLKEA 59
DB 1367 ECNPKES---YPMDCNKNIDISHD-GACMPRRKLCVYIAHSQENIKTDDNLKDA 1422
QY 60 FIKSAAEFFSWYKSK-DGEGNELDKELKGRIPAPLRSMEYTGDFYDFLFGDI 118
DB 1423 FIKTAAEFFLSWQYKSKNDSEAKILDR---GLIPQFLASMAVTFGDRDCLMTDI 1478
QY 119 SKHGSGSLKQIDSLFKNQDQSPNGKTRQEWMTESHSIHWAMCALYKIGAKKD- 177
DB 1479 SKQNDVAAKADKIGKFFSKDSKSPGSLRQEWMTKNGPEIWMGMLCALYKVTDTDNK 1538
QY 178 --FTENYGNVNFSDK-STLLEFAKRPQFLRWLTWYDDCYTRQKYLKDVQEKCS- 233
DB 1539 RKINDYSYDKVNOSONGPNLSLEFPAKQFLRWMIEMGEFCERKKNIIKDACNEI 1598
QY 234 NDOLKCDT---ECNKKCEDYKVKM-KRKEWI-----PODKYYK 268
DB 1599 NSTQOCNDAKRCNQACRAIOEYENKKEFSQTNNEVLKANYQPDPEYK 1650
```

RESULT 3

US-09-210-288-12
Sequence 12, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-12

Query Match 40.4%; Score 624; DB 4; Length 2710;
Best Local Similarity 43.8%; Pred. No. 8, 5e-53;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPRKNSGYDPMOC-GNINIVEDPRVCMPPRRKLCYHFLANDNEIKKLSQVNLKKA 59
DB 1367 ECHPKES---YPMDCNNIDISHD-GACMPPRKLCLYIAHESQTEIKITDDMLKDA 1422
QY 60 FIKSAAEFFSWYXXSKR-DGEGNLDKELKRGKIPPAFLSMFTFGDYRDLFGTDI 118
DB 1423 FIKTAAEFELSMQYXXSKNDSEAKILDR---GLIPSGFLRSMWMTTFGDRICLNTDI 1478
QY 119 SKHGEGSKLKEQIDSLFNGDOKSPNGKTRQEWTEHSHETIEMALCALVTKGAKKD- 177
DB 1479 SKQNDVAAKAKDKIGKFEFSKDSKSPGSLRQEWMTNNGPEIKGMICALTKYVTDI 1538
QY 178 --FTENYGNVNFSDK-STTLEEFARPOFLRLTEWTDYCYTROKYLKDVQEKCS- 233
DB 1539 RKIKNDYSTDKVNOSONGPNLEEFAPAKPOFLKMWLEWGEFECAEKOKENIITKDCNET 1598
DB 234 NDOLKCDT---ECNKKCEDYVYKMK-KKKEWI-----PODKYRK 268
DB 1599 NSTQOCNDAKHRCNOACRAVQEVYENKKKEFSQGTNNFVLKANVQPDPEYK 1650

RESULT 4

US-08-487-826B-14
Sequence 14, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMIDIUM VIVAX
TITLE OF INVENTION: AND PLASMIDIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487, 826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29, 655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3060 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-826B-14

Query Match 40.4%; Score 624; DB 2; Length 3060;

Best Local Similarity 43.8%; Pred. No. 1e-52;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPRKNSGYDPMOC-GNINIVEDPRVCMPPRRKLCYHFLANDNEIKKLSQVNLKKA 59
DB 1365 ECHPKES---YPMDCNNIDISHD-GACMPPRKLCLYIAHESQTEIKITDDMLKDA 1420
QY 60 FIKSAAEFFSWYXXSKR-DGEGNLDKELKRGKIPPAFLSMFTFGDYRDLFGTDI 118
DB 1421 FIKTAAEFELSMQYXXSKNDSEAKILDR---GLIPSGFLRSMWMTTFGDRICLNTDI 1476
QY 119 SKHGEGSKLKEQIDSLFNGDOKSPNGKTRQEWTEHSHETIEMALCALVTKGAKKD- 177
DB 1477 SKQNDVAAKAKDKIGKFEFSKDSKSPGSLRQEWMTNNGPEIKGMICALTKYVTDI 1536
QY 178 --FTENYGNVNFSDK-STTLEEFARPOFLRLTEWTDYCYTROKYLKDVQEKCS- 233
DB 1537 RKIKNDYSTDKVNOSONGPNLEEFAPAKPOFLKMWLEWGEFECAEKOKENIITKDCNET 1596
DB 234 NDOLKCDT---ECNKKCEDYVYKMK-KKKEWI-----PODKYRK 268
DB 1597 NSTQOCNDAKHRCNOACRAVQEVYENKKKEFSQGTNNFVLKANVQPDPEYK 1648

RESULT 5

US-08-568-459A-8
Sequence 8, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMIDIUM VIVAX
TITLE OF INVENTION: AND PLASMIDIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568, 459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29, 655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
MOLECULE TYPE: linear
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-8

Query Match 32.0%; Score 493.5; DB 2; Length 921;
Best Local Similarity 40.0%; Pred. No. 1.8e-40;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPKNSNGYDPQMGGINL-VEPRVCMPPRRQKLCYHFLANDNEIKLQSOVN-LKE 58
DB 414 ECRKFT---ISEWTCDESKIKMGHACIPPRQKLCIHL-----EKIMTNNELKY 464
QY 59 AFKSAAEFFSWY-YKSKDGESELDKELKSGKIPAPLRSMFYFGDYRDLFGTD 117
DB 465 AFKCAAEFTFLWQNYKKDKNGNAEDDEKLGITIPEDFRQMFYFFADYRDLCLGTD 524
QY 118 I-----SKGHEGSKLEQDLSLFKNGDQSPNG-KTROEMTSHSEIWEAMICAL-- 168
DB 525 ISSKDTSGVG---KVKCNIDVF---YKISNIRYRKSMWETNGPVIMEGMICALSY 577
QY 169 -----VKIGAKKDDFTENYGYNNYK-SDKSTLEEFARPOFLRWLTWYDYCYTRQ 221
DB 578 DTSLNVPNPEHKKLTGNNNEKVIIFGSDSSTILSKFSERPOLRWLTWGENFCEOK 637
QY 222 KYLDVQEKCKSNQDKDTECNK---CEDYK-----YMKKKREWIPODK- 265
DB 638 KEYVLLAKCKDCD-VDGDCKNGCKVACDKQCKOYHSHWIGIWDNYKKQGRYTEVKKI 696
QY 266 -YKDERDK 274
DB 697 PLYKEDKDVK 706

RESULT 6
US-08-826B-8
Sequence 8, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-826B-8

Query Match 32.0%; Score 493.5; DB 2; Length 921;
Best Local Similarity 40.0%; Pred. No. 1.8e-40;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPKNSNGYDPQMGGINL-VEPRVCMPPRRQKLCYHFLANDNEIKLQSOVN-LKE 58
DB 414 ECRKFT---ISEWTCDESKIKMGHACIPPRQKLCIHL-----EKIMTNNELKY 464
QY 59 AFKSAAEFFSWY-YKSKDGESELDKELKSGKIPAPLRSMFYFGDYRDLFGTD 117
DB 465 AFKCAAEFTFLWQNYKKDKNGNAEDDEKLGITIPEDFRQMFYFFADYRDLCLGTD 524
QY 118 I-----SKGHEGSKLEQDLSLFKNGDQSPNG-KTROEMTSHSEIWEAMICAL-- 168
DB 525 ISSKDTSGVG---KVKCNIDVF---YKISNIRYRKSMWETNGPVIMEGMICALSY 577
QY 169 -----VKIGAKKDDFTENYGYNNYK-SDKSTLEEFARPOFLRWLTWYDYCYTRQ 221
DB 578 DTSLNVPNPEHKKLTGNNNEKVIIFGSDSSTILSKFSERPOLRWLTWGENFCEOK 637
QY 222 KYLDVQEKCKSNQDKDTECNK---CEDYK-----YMKKKREWIPODK- 265
DB 638 KEYVLLAKCKDCD-VDGDCKNGCKVACDKQCKOYHSHWIGIWDNYKKQGRYTEVKKI 696
QY 266 -YKDERDK 274
DB 697 PLYKEDKDVK 706

RESULT 7
US-09-210-288-8
Sequence 8, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobb Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1PMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-0176
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:

LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-8

Query Match 32.0%; Score 493.5; DB 4; Length 921;
Best Local Similarity 40.0%; Pred. No. 1.8e-40;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPKNSNGYPMOCGNINL-VEPRVCMPPRRKICVHFLAANDIEKKLSQVN-LKE 58
DB 414 ECRKRT---YSEWTCDESKIKMGHACIPRRKICLHYL-----EKIMNTLEKY 464
QY 59 APTKSAAEFTFSWY-YYSKDGNEELDKKEGKIPPAFLRSMTYTGDDYDFLFGTD 117
465 APTKSAAEFTFLMONTYKDKNNAEDLEKLGIGIPEDFKROMFTYFADYDIDLGTD 524
QY 118 I-----SGHGSGKLEKQIDSLFKNGDQKSPNG-KTQEWMTTEHSHEIWEAMLCAL-- 168
DB 525 ISKKTPTSGVG---KVCNIDVF---YKISNSIRKRSWMTETGPIWESMLCALST 577
QY 169 -----VKIGAKKDPFTENYGYNNYKF-SDKSTTLEFAKRPQFLMLTEWYDYCYTRQ 221
DB 578 DTSLNVPNETHKKLEGNNNPEKVFEGSDSSTLTSKFSERPQFLMLTEWNGSNFCKEAK 637
QY 222 KTKLDYQEKCKSDQLKCTEKNK---CEDYVK-----YMKKKEMIPQDK- 265
DB 638 KEKTVLACKCKDCD-VDGSGKCKGKCVACKDCKQYHSMWIGIWDNYKKOKGRYEVKKI 696
QY 266 -YKKEDRDK 274
DB 697 PLYKEDKDYK 706

RESULT 8

US-08-487-826B-16
Sequence 16, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487, 826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2182 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
TYPE: amino acid
ANTISENSE: NO
FRAGMENT TYPE: N-terminal
US-08-487-826B-16

Query Match 24.8%; Score 382.5; DB 2; Length 2182;
Best Local Similarity 28.0%; Pred. No. 6.4e-29;
Matches 115; Conservative 50; Mismatches 100; Indels 145; Gaps 18;

QY 2 CHPKNSNG---YPMOCGNINL-----EDPRVCMPPRRKICVHFLA-----NDNEI 47
DB 869 CGLKYPGGKEKPPNMKCVTPSGVSTATSGKDALICVPPRRRLYVGGLSQMSRGDEI 928
QY 48 KTIQSOV-----NKEAFKSAAEFTFSWYTK-----SKDG----- 80
DB 929 TEVSSSEATSNAPSGSESEKLTAFIESALETFLMHRKYEEKKPPATODGAGIGVSLPEP 988
QY 81 --EGNELELKE-GKIPPAFLRSMTYTGDDYDFLFGTDISKHG----- 123
DB 989 SPGEEDPOTQLOQTGYIPDFLROMFTYTLADYDILYSGSNDTSDTTGKOTPESSNDNIK 1048
QY 124 -----EGSKLEKQIDSLFKNGDQKSPNGK-----ROEWTEHS 157
DB 1049 NIYLAASGSTEKEKMK-QIQAKIK-----KILNGATSGVPYTKNSVTPQOTWMENTA 1103
QY 158 HEIWEAMLCAL-----VKI-----GAKKDPFTENYGYNNYKESDKS- 193
DB 1104 KDIWNAVICALTYKENDARGTSAKIEONKDLKALWDEANKNPRIKYOYTNKLEDESG 1163
QY 194 -----TTLEFAKRPQFLMLTEWYDYCYTRKTYKLDYQEKCKSDQLK----- 238
DB 1164 AKSNDTIQPTTLKNEVEIIPFFRWLHEWNGSFFFEAKRIQAQIKHCDMDGCKOYSGDG 1223
QY 239 --CD-----TECNKCKEDYVYMKKK-EMIPQDKYVYKDER 271
DB 1224 EYCEEIYSQYNYLQDLSSSCAKPCRLYKTKWIEKKKTEYKQKAAVEQK 1273

RESULT 9

US-08-568-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 17.8%; Score 274; DB 2; Length 1435;
Best Local Similarity 29.2%; Pred. No. 2.2e-18;
Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;

QY 14 WCCGNINLVEDPRVCMPPRRQKLCVHFLANDNEIKKIQSOVNLKEAFIKSAAEFTFSWY 73
DB 474 WECKNPYLSTKDYCVPPRRQELC---LGNIDRIYD-KNLMITEHLAIAIESILKR 529
QY 74 YKSKDQSGNLDKELKGIKIPAPLASMFTYEDYDPLFGDIDSKGHEGSKLEQID 133
DB 530 KYKNKD-----DKEV-----CKIINKTFADIRDIIGTDYNDL-SNRKLVGKIN 573
QY 134 SLKNDQKSPNGKT-ROEWTEHSEIWEAMCALVKGAKKDDFTENGYNNVFSK 192
DB 574 TNSKYVHRNKKNDKLFDEWKKVYKKNVNI-----SWFKDK 612
QY 193 STLEEFAR-POPLRMLEWYDYCYTRQKYLKDVQEKCKSNDOLKC-DTECNKCEYD 250
DB 613 TVCKEDDIENTIPQFFNFSEMGDDYCODKTKMETILKVECK---EKPCDDNCKSKCNSY 669
QY 251 VKYM-KKKKEMIPQDKYKD 269
DB 670 KEMISKKEEYNNQAKOYOE 689

RESULT 10
US-08-487-826B-4

Sequence 4, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-4

Query Match 17.8%; Score 274; DB 2; Length 1435;
Best Local Similarity 29.2%; Pred. No. 2.2e-18;
Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;

QY 14 WCCGNINLVEDPRVCMPPRRQKLCVHFLANDNEIKKIQSOVNLKEAFIKSAAEFTFSWY 73
DB 474 WECKNPYLSTKDYCVPPRRQELC---LGNIDRIYD-KNLMITEHLAIAIESILKR 529
QY 74 YKSKDQSGNLDKELKGIKIPAPLASMFTYEDYDPLFGDIDSKGHEGSKLEQID 133
DB 530 KYKNKD-----DKEV-----CKIINKTFADIRDIIGTDYNDL-SNRKLVGKIN 573
QY 134 SLKNDQKSPNGKT-ROEWTEHSEIWEAMCALVKGAKKDDFTENGYNNVFSK 192
DB 574 TNSKYVHRNKKNDKLFDEWKKVYKKNVNI-----SWFKDK 612
QY 193 STLEEFAR-POPLRMLEWYDYCYTRQKYLKDVQEKCKSNDOLKC-DTECNKCEYD 250
DB 613 TVCKEDDIENTIPQFFNFSEMGDDYCODKTKMETILKVECK---EKPCDDNCKSKCNSY 669
QY 251 VKYM-KKKKEMIPQDKYKD 269
DB 670 KEMISKKEEYNNQAKOYOE 689

RESULT 11
US-09-210-288-4

Sequence 4, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
CORRESPONDENCE ADDRESS:
ADDRESS: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-4

Query Match 17.8%; Score 274; DB 4; Length 1435;
Best Local Similarity 29.2%; Pred. No. 2,2e-18;
Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;

QY 14 WCGNINLVDPVPCPPRRQKICVHFLANDNEIKLOSQVNLKEAFISAAETFSWYKSGEGNELDK 73
DB 474 WEKNYIISTKDYCVPPRROELC--LGNIDRIYD-KNLMKEHILAIATIESRILKR 529
QY 74 YKSKDEGENELDKELKEGKIPAPLRSMFYTGDRDLEFGDISKSGEGSKLEKQID 133
DB 530 KYTKND-----DKEY-----CKITKTPADRIDIIIGTDYNNDL-SNRKLVGKIN 573
QY 134 SLEKNDQKSPNGKT-RQSWTEHSHIEWAMLCALVKGAKKDDFTENGYNNVFSDK 192
DB 574 TNSKYHRKKNDKLFREDWMVKYIKKDVNVI-----SWFKKK 612
QY 193 STLEEFAR-POFLWLTLEWIDDYCYTRQKYLKDVQEKSKSDQKLC-DTECNKCEDY 250
DB 613 TYCKEDDINIEIPQFFRMFSEWGDYCODYTKMLETILKVECK--EKPCEDDNCKSKCNSY 669
QY 251 VKYM-KKKKEMIPQDKYKD 269
670 KEWISKKEEYKQAKOYOE 689

RESULT 12
US-08-568-459A-21
Sequence 21, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-568-459A-21

Query Match 17.7%; Score 273.5; DB 2; Length 311;
Best Local Similarity 27.5%; Pred. No. 3e-15;
Matches 66; Conservative 22; Mismatches 141; Indels 11; Gaps 4;

QY 28 CMPRRQKICVHFLANDNEIKLOSQVNLKEAFISAAETFSWYKSGEGNELDK 87
DB 24 CMPRRQKICLVYIXX 83
QY 88 ELKEGKIPAPLRSMFYTGDRDLEFGDISKSGEGSKLEKQIDSLFKNDQKSPNGK 147
DB 84 XXXXXX--QFLRSMYTGDRDICTLNDISFKQNDVXXXXXXXXXXXXXXXXX 140
QY 148 TROEMWTEHSHIEWAMLCAL--VKIGAKKDDFTENGYNNVFKSDKSTLEEFAR 203
DB 141 SROEMWKTNGPELTWKGCLLXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 200
QY 204 QFLRWLTLEWIDDYCYTRQKYLKDVQEK--KSNQDLKCDT--DCNKKCEDYVYMKKKKE 259
DB 201 QFLRWLTLEWIDDYCYTRQKYLKDVQEK--KSNQDLKCDT--DCNKKCEDYVYMKKKKE 260

RESULT 13
US-08-487-826B-33
Sequence 33, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible

```

1 CURRENT APPLICATION DATA:
2 APPLICATION NUMBER: US-09/210,288
3 FILING DATE:
4 CLASSIFICATION:
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Fuller, Michael
7 REGISTRATION NUMBER: 36,516
8 REFERENCE/DOCKET NUMBER: NIH121.1EWdVdY
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (619) 235-8550
11 TELEFAX: (619) 235-0176
12 INFORMATION FOR SEQ. ID NO.: 21:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 311 amino acids
15 TYPE: amino acid
16 STRANDEDNESS: single
17 TOPOLOGY: linear
18 MOLECULE TYPE: peptide
19 HYPOTHEetical: NO
20 AMTI-SENSE: NO
21 FRAGMENT TYPE: Internal
22 ORIGINAL SOURCE:
23 US-09-210-288-21

```

Query Match	17.7%	Score 273.5	DB 4	Length 311
Best Local Similarity	27.5%	Pred. No. 3e-19		
Matches	66	Conservative	22	Mismatches 141, Indels 11, Gaps 4

QY	28	CMPPRRKCLCVHFLANDNEIKRLQSGVNLKPAFTKSAALTFEFSWYKKRSGDEGNELDK	87
Db	24	CMPPRRKCLCVYIYXX	83
QY	88	ELKEGRIIPAFRSMFYTFGRDRLFTGDISKGGESKIKLEIDSLFNGQKSPNGK	147
Db	84	XXXXXX--QFLRSMATYTFGRDRLCLMTDISKQNDVXXXXXXXXXXXXXXXXXSKSPSGI	140
QY	148	TROEMWTEHSHEIWEAMCAL---VKIGAKDDFTENYGYNNVFKSDKSTYLEEFAKRP	203
		: : :	
Db	141	SRQEMWTKGAPETIMKGMICALXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXP	200
QY	204	QFLRMILTEWYDDCYTRQKYLKDVQERC-KSNDOLKCDT---ECNKCEDYVYIKMKKE	259
		: : : :	
Db	201	QFLRMILTEWEEECAROKKENIKIDACKXXXXXXXXXXCKXKHKHRCNQACRAVGYGVNKKK	260

RESULT 15
US-08-568-459A-6
; Sequence 6, Application US/08568459A

; Patent NO. 5849306
; GENERAL INFORMATION:

1 APPLICANT: Miller, Louis H.
 2 APPLICANT: Peterson, David S.
 3 APPLICANT: Su, Xin-zhaun.
 4 APPLICANT: Wellens, Thomas E.
 5 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 6 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 7 NUMBER OF SEQUENCES: 37
 8 CORRESPONDENCE ADDRESSES:
 9 ADDRESSEE: Knobbe Martens Olson & Bear
 10 STREET: 620 Newport Center Drive 16th Floor
 11 CITY: Newport Beach
 12 STATE: California

```

? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.25
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/568,459A

```

FILED DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned
REGISTRATION NUMBER: 29, 655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
JIS-08-568-459A-6

Query Match 16.9%; Score 260; DB 2; Length 749;
Best Local Similarity 28.7%; Pred. No. 2, 2e-17;
Matches 78; Conservative 41; Mismatches 93; Indels 60; Gaps 13;

QY 14 WCGNINLYEDPR-VCMPPRQKLVH--FLANDNEIKKIQSOVNLKEAFIKSAAEFF 70
DB 125 WNCYSNNKTYKPEGVCGPPRQQLCGYIFLIRGNEEGIKDHIIN-----KAANYEAMH 178
QY 71 SWYYKSKDQEGNELDKELKEGKIPAPLRSMFYTFGRDPLFGTIDISKGHGEGSKLKE 130
DB 179 LKEKYENAGD-----KICNAIILGS-----YADIGDIYRGLDVWRDINT--NKLSE 222
QY 131 QIDSLFKNGD--QKSPNGKTRQEWNTESHSEIWEAMLCALYKIGAKKDDFTENGYNNY 187
DB 223 KFKIKFMGGSNSRKKQNDNNERNKMKWKKQKRNLIWSSM-----Y 260
QY 188 KPSDKSTLE--EFAKRPQFLRWLTWTDDYCYTRQKTIKADYQEKCKSNDQLKC--DTEC 243
DB 261 KHLPKGKTCKRHNNFEKIPQFLRWLKEWGDCECEMGTEYKOLEKICENKN--CSEKKC 317
QY 244 NKKCEDYVYKMKKK--EWIPQDKYKDERDKK 274
DB 318 KNAKCSYEKWKERKNEYNLQSK--KFDSDK 347

Search completed: April 28, 2003, 10:32:11
Job time : 7.78453 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:32:15 ; Search time 6.16575 Seconds

(without alignments)
3586.886 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 1543
Sequence: 1 DCHPRKNSNGYPMQCGNIN.....KKRWIPDKYKDEKRRK 276Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

Optimal number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	624	40.4	2710	US-10-153-273-12	Sequence 12, Appl
2	493.5	32.0	921	US-10-153-273-8	Sequence 8, Appl
3	289.5	18.8	1143	US-09-924-154-14	Sequence 16, Appl
4	282	18.3	972	US-09-924-154-16	Sequence 14, Appl
5	274	17.8	1435	US-10-153-273-4	Sequence 4, Appl
6	273.5	17.7	311	US-10-153-273-21	Sequence 21, Appl
7	261	16.9	1421	US-09-924-154-13	Sequence 13, Appl
8	260	16.9	749	US-10-153-273-6	Sequence 6, Appl
9	260	16.9	1086	US-09-924-154-15	Sequence 15, Appl
10	223	14.5	1115	US-10-153-273-2	Sequence 2, Appl
11	218.5	14.2	324	US-10-153-273-17	Sequence 17, Appl
12	210	13.6	700	US-10-153-273-10	Sequence 10, Appl
13	209	13.5	1501	US-09-924-154-17	Sequence 17, Appl
14	164	10.6	277	US-10-153-273-15	Sequence 15, Appl
15	142	9.2	282	US-10-153-273-16	Sequence 16, Appl
16	140	9.1	291	US-10-153-273-13	Sequence 13, Appl
17	133.5	8.7	411	US-10-153-273-19	Sequence 19, Appl
18	126	8.2	411	US-10-153-273-20	Sequence 20, Appl
19	123	8.0	362	US-10-153-273-18	Sequence 18, Appl

20	108	7.0	271	9	US-10-153-273-14	Sequence 14, Appl
21	91	5.9	351	10	US-09-864-761-33332	Sequence 33332, A
22	90	5.8	2332	9	US-09-957-641-2	Sequence 2, Appl
23	90	5.8	2332	9	US-10-187-319-2	Sequence 2, Appl
24	90	5.8	2351	9	US-10-132-829-4	Sequence 4, Appl
25	89	5.8	410	10	US-09-815-242-13393	Sequence 13393, A
26	89	5.8	410	10	US-09-815-242-13628	Sequence 13628, A
27	87	5.6	793	10	US-09-900-237-18	Sequence 18, Appl
28	86	5.6	1139	9	US-09-820-843A-15	Sequence 15, Appl
29	85.5	5.5	1805	9	US-09-820-843A-73	Sequence 73, Appl
30	85	5.5	632	9	US-10-108-605-287	Sequence 287, App
31	85	5.5	956	9	US-10-000-256A-153	Sequence 153, App
32	85	5.5	1261	12	US-10-147-268-2	Sequence 2, Appl
33	84.5	5.5	701	9	US-09-991-496-16	Sequence 16, Appl
34	84.5	5.5	701	10	US-09-874-923-16	Sequence 16, Appl
35	84	5.4	199	9	US-09-864-761-34284	Sequence 34284, A
36	84	5.4	711	9	US-09-298-523B-3	Sequence 3, Appl
37	82.5	5.3	286	9	US-10-106-534-4	Sequence 4, Appl
38	82.5	5.3	503	9	US-09-738-626-5485	Sequence 5485, Ap
39	80	5.2	589	9	US-09-298-523B-14	Sequence 14, Appl
40	80	5.2	690	9	US-09-298-523B-61	Sequence 61, Appl
41	80	5.2	691	9	US-09-298-523B-1	Sequence 1, Appl
42	80	5.2	701	9	US-09-298-523B-62	Sequence 62, Appl
43	80	5.2	905	9	US-10-114-893-127	Sequence 127, App
44	80	5.2	915	10	US-09-880-192-359	Sequence 59, Appl
45	80	5.2	929	9	US-09-298-523B-60	Sequence 60, Appl

ALIGNMENTS

RESULT 1
US-10-153-273-12
Sequence 12, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhan
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21 May 2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1.FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 12;
US-10-153-273-12

Query Match 40.4%; Score 624; DB 9; Length 2710;
Best Local Similarity 43.8%; Pred. No. 3.2e-46;
Matches 128; Conservative 49; Mismatches 83; Indels 32; Gaps 11;

QY 1 DCHPKRNSNGYPMQC-GNINLVEDPRVCMPRROROKLCVHFANDNEIKKLOSOVNLKE 59
DB 1367 ECPNKEKES---YPMDCNNIDISHD-GACMPRROROKLCVYINHEQOTENIKTDMLKDA 1422
60 FIKSAAEFFEFYKSK-DGEGNELDEKELKGIIPAFLSMFTYFGDYRDLFGTDI 118
1423 FIKSAAEFFELSMOYKSKNDSEAKTILDR---GLIPSOFLRSMFTYFGDYRDLFGTDI 1478
QY 119 SKGHGSKLEQIDSLFKNGDQSPNGKTRQEWTEHSHETMEALCALYKGAKKD- 177
DB 1479 SKKQNDYAKAKDKIGFEKDGSKSPSGLSRQEMWKTNGPELWKMICALYKVTDTDNK 1538
QY 178 --FTETYGNNVYKESDK-STTLEEFARPOPLRWLTETWYDYCYTQKYLKDVQEKCS- 233
DB 1539 RKIKNDYSYDKVQSONGPNSELEFAKPOFLRMMEWEEBCECAEKKENIKKDCNFI 1598
QY 234 NQJLKCDT---EONKCEYVYKMK-KKEWT-----PODKTYK 268
DB 1599 NSTQOCNDAKHCNACRAVYOEYVENKKKEFGQTNNFVLKANVOPDPEYK 1650

RESULT 2

US-10-153-273-8
Sequence 8, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

NUMBER OF SEQUENCES: 37 AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153/273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121,1FMDV1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:

ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 8;
US-10-153-273-8

Query Match 32.0%; Score 493.5; DB 9; Length 921;
Best Local Similarity 40.0%; Pred. No. 2.8e-35;
Matches 124; Conservative 39; Mismatches 94; Indels 53; Gaps 15;

QY 1 DCHPKRNSNGYPMQC-GNINLVEDPRVCMPRROROKLCVHFANDNEIKKLOSOVNLKE 58
DB 414 ECTKRET---YSEWDCESKIMGQACIPRROROKLHYL-----EKTMTNTEIKY 464
QY 59 AFIKSAAEFFEFYKSK-YKSKDGEENELDEKELKGIIPAFLSMFTYFGDYRDLFGTD 117
DB 465 AFIKSAAEFFELSMOYKSKNDSEAKTILDR---GLIPSOFLRSMFTYFGDYRDLFGTD 524
QY 118 I-----SKGHGSKLEQIDSLFKNGDQSPNG-KTROEWTEHSHETMEALCALY- 168
DB 525 ISSKDTSKRGV---KVCNIDIVF---YKINSIRYKFSMWEINGVINEGMICALSY 577
QY 169 -----YVIGAKKDFTENYGNVNF--SDKSTTLEEFARPOPLRWLTETWYDYCYTRQ 221
DB 578 DTSLNANVPETHKLTBEGNNNEKVIYFSDSTYLSKFSERPOPLRWLTETWYDYCYTRQ 637
QY 222 KYLKDVQEKCSNDQKCDTECNK---CEDYK-----YKKKKKEWIPDK- 265
DB 638 KEYKVLAKCKDQD-VDDGKCGKGVACQKQCHYHISGWIWIDNKKKQGRTEYK 696
QY 266 -YYNDERDK 274
DB 697 PLYKEDKVK 706

RESULT 3

US-09-924-154-14
Sequence 14, Application US/09924154
Patent No. US20020127241A1

GENERAL INFORMATION:

APPLICANT: Natum, David L.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 14

LENGTH: 1143

TYPE: PRT

ORGANISM: Mammalian
US-09-924-154-14

Query Match 18.8%; Score 289.5; DB 10; Length 1143;
Best Local Similarity 28.4%; Pred. No. 3.5e-17;
Matches 80; Conservative 45; Mismatches 100; Indels 57; Gaps 13;

QY 2 CHPKRNSNGYPMQC-GNINLVEDPRVCMPRROROKLCV---HFLANDNEIKKLOSOVNLKE 58
DB 462 CNK--SNKSNMNCIGTFNKKFGTCPEPRQTLCLGRTYLLHRGHE-----EDYKE 512

QY 59 APTKSAAEFTFSWYKSKDGEKGNELDKELKGIIPAFLENSMYTFGDRPLEGTDI 118
DB 513 HILGASITPAQLLKTKYKKEKEN-----ALCSITONSADLADIINGSDI 557
QY 119 SKGHGSGSLKQIDSLFNGDOKSPNGKT-ROEWTEHSHEIWEAMLCALVIGAKKDD 177
DB 558 IKDY-YGKKEEMLNKNVNDKRNNEESLKIIFREKWDENKENVKMSAVL-----KNKE 611
QY 178 FTENGYNNVNFSDKSTLEEFKARPOFLRWLTLEYDYCYTROKYL---KDVQEKCSN 234
DB 612 TCKDY-----DKFOKIPQFLRWLFREKMGDDCEKREKIYSPESKVECKK 657
QY 235 DOLKCD-TECNKKCEDYVYKMK-KKKEWIPQ-DKYVXDERDK 273
DB 658 D--CDENTCKNCKSEYKKWIDLKSEYKQVDKTKDKKKK 696

RESULT 4
US-09-924-154-16
Sequence 16, Application US/09924154
Patent No. US20020127241A1
GENERAL INFORMATION:
APPLICANT: Natum, David L.
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
CURRENT FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn version 3.1
SEQ ID NO 16
LENGTH: 972
TYPE: PRT
ORGANISM: Mammalian
US-09-924-154-16

Query Match 18.3%; Score 282; DB 10; Length 972;
Best Local Similarity 27.8%; Pred. No. 1,3e-16;
Matches 84; Conservative 45; Mismatches 105; Indels 68; Gaps 12;

QY 2 CHPKNSNGYPMQGNINLVEDPRVCMPPRROKLCVHFLAND-----NEIKKLQSOVNL 56
DB 138 CEKGNDR--TWQCINEHIKDPDVCPPRROQLCLGNLDKDEFKVNLDLKKFLNRI-- 192
QY 57 KEAFIKSAAEFTFSWYKSKDGEKGNELDKELKGIIPAFLENSMYTFGDRPLEGT 116
DB 193 ----ILGIDEGKFLLEKTKRNHNNMYLDERA-----CKYLNYSDDYKNITILK 239
QY 117 D-----ISKGHSGSLKQIDSLFNGDOKSPNGKTROEWTEHSHEIWEAM 164
DB 240 DMMRDNSIKTENILKGNFEG--IKANYISMYSYADLSID-EFKHMDQNKQJLEAI 236
QY 165 LCALVIGAKKDDFTENGYNNVNFSDKSTLEEFKARPOFLRWLTLEYDYCYTROKYL 224
DB 297 SC-----EFGKGNHT-----GVCLEMDNDNDQYILHMFREKMKDFCIDILKMN 338
QY 225 KDVQEK-----KSDQLKCDTECNKKCEDYVYK--KKKKEWIPQDKYKDERDK 274
DB 339 DYKKEPCIDKVKKSPRSENPSDVATVCKNSCTDYDKMIINRKREYKMOSSKYK--RDRS 396
QY 275 RF 276
DB 397 LF 398

RESULT 5
US-10-153-273-4
Sequence 4, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-153-273-4

Query Match 17.8%; Score 274; DB 9; Length 1435;
Best Local Similarity 29.2%; Pred. No. 1.1e-15;
Matches 76; Conservative 43; Mismatches 93; Indels 48; Gaps 11;

QY 14 WQGNINLVDPVCMPPRROKLCVHFLANDNEIKKLQSOVNLKEAFIKSAAEFTFSWY 73
DB 474 WECKNPILSTKDYCVPRROELC--LGNIDRIYD-KNLMKKEHLAIAIYESRLIKR 529
QY 74 YKSKDGEKGNELDKELKGIIPAFLENSMYTFGDRPLEGTDISKGHSGSLKQID 133
DB 530 KYKND-----DKV-----CKIINKTRADIDITIGGIDYNDL-SNRKLVGKIN 573
QY 134 SLFNGDOKSPNGKT-ROEWTEHSHEIWEAMLCALVIGAKKDDFTENGYNNVNFSDK 192
DB 574 TNSKYVHRNKNKLFEDKEMKVIKKDVMYV-----SWYKDK 612
QY 193 STLEEFKAR-POFLRWLTLEYDYCYTROKYLKDVQEKSKDQKLC-DTECNKKCEDY 230
DB 613 TVCKEDDIENIPQFPRFWSMGDDYCODKTKMTETLKVCK--EKPCEDNCKSKNSY 669
QY 251 VKYM-KKKKEWIPQDKYKDD 269
DB 670 KEMISKKEEYNNQAKQYOYE 689

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-153-273-6

Query Match 16.9%; Score 260; DB 9; Length 749;
Best Local Similarity 28.7%; Pred. No. 8.5e-15;
Matches 78; Conservative 41; Mismatches 93; Indels 60; Gaps 13;

QY 14 WCGNINLVEDPR-VCMPPRRQKICVH--FLANDNEIKILOSQVNLKEAFKSAATFF 70
DB 125 WNCYSNNKVKPRGVCPPRQOLCLGYIFLRDGNBGLKDHIN-----KAANYEAMH 178
QY 71 SWYYSKQEGNDELKELKEGKIPAFILSMFYTGDIYDFLFGTIDISKHGBGSLKE 130
DB 179 LKEYEYAGD-----KICNAIIGS---YADIGDIVAGLDVWDHINT-NKLSE 222
QY 131 QIDSLFRNGD---OKSPNGKTRQEWTEHSHIWEAMICALVKGAKKDDFTENYGYNNV 187
DB 223 KFKITWGGNSRKKNQDNNEKRWKWKORNLWSS-----V 260
QY 188 KFSKSTTLE--EFAKRPQFLRWLTWYDYCYTRKYLKDVQCKCKSNDQKLC-DTEC 243
DB 261 KHIPKGTCKRHNNFEKIPQFLRWLKEWGEFCEMGTEVQLEKICENKN---CSEKKC 317
244 NKCEDEVKYMKKK-EMIPQDKYKDERDK 274
DB 318 KNACSSYEKWKERKNEYNLQSK--KFSDSK 347

RESULT 9
US-09-924-154-15
Sequence 15, Application US/09924154
Patent No. US20020127241A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
FILE REFERENCE: 05213-0465 43170-262105
CURRENT APPLICATION NUMBER: US/09/924,154
PRIOR FILING DATE: 2001-08-07
PRIOR APPLICATION NUMBER: US 60/223,525
PRIOR FILING DATE: 2000-08-07
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin version 3.1
SEQ ID NO 15
LENGTH: 1086
TYPE: PRT
ORGANISM: Mammalian
US-09-924-154-15

Query Match 16.9%; Score 260; DB 10; Length 1086;
Best Local Similarity 28.7%; Pred. No. 1.3e-14;
Matches 78; Conservative 41; Mismatches 93; Indels 60; Gaps 13;

QY 14 WCGNINLVEDPR-VCMPPRRQKICVH--FLANDNEIKILOSQVNLKEAFKSAATFF 70
DB 367 WNCYSNNKVKPRGVCPPRQOLCLGYIFLRDGNBGLKDHIN-----KAANYEAMH 420
QY 71 SWYYSKQEGNDELKELKEGKIPAFILSMFYTGDIYDFLFGTIDISKHGBGSLKE 130
DB 421 LKEYEYAGD-----KICNAIIGS---YADIGDIVAGLDVWDHINT-NKLSE 464
QY 131 QIDSLFRNGD---OKSPNGKTRQEWTEHSHIWEAMICALVKGAKKDDFTENYGYNNV 187
DB 465 KFKITWGGNSRKKNQDNNEKRWKWKORNLWSS-----V 502
QY 188 KFSKSTTLE--EFAKRPQFLRWLTWYDYCYTRKYLKDVQCKCKSNDQKLC-DTEC 243
DB 503 KHIPKGTCKRHNNFEKIPQFLRWLKEWGEFCEMGTEVQLEKICENKN---CSEKKC 559
QY 244 NKCEDEVKYMKKK-EMIPQDKYKDERDK 274
DB 560 KNACSSYEKWKERKNEYNLQSK--KFSDSK 589

RESULT 10
US-10-153-273-2
Sequence 2, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH21.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1115 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium vivax
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-153-273-2

Query Match 14.5%; Score 223; DB 9; Length 1115;
Best Local Similarity 26.1%; Pred. No. 2.5e-11;

Matches 70; Conservative 44; Mismatches 90; Indels 64; Gaps 14;

QY 13 DWOCGNINLVEDPRVCMPRRKCLVHFLAN-----DNEIKLOSQVNLKEAFI 61
DB 272 DWDC---NFKRD--VCIDRRYQCLMKELINLVNNTDTHFRDITFRKLY---LKRRLI 322
QY 62 KSAAEFTFSYVYKSKDGEENELDKELKEGIPPAFLRSMEYFGDYRDFLFGTDISK 121
DB 333 YDAVE-----GDLLKLNRYRNKDFCKDIRMSLGPFGDIMGTDN-EG 366
QY 122 HEGSKLKEQIDSLFKNGDOKSPNGKROEWTSHSEIMEAMICALVKIGAKKDPFTEN 181
DB 367 IGYSKVVENNRSLF-GTDEKAQ--QRRKQWNSKAQIWTAMMYSVK--RLKGNFTWI 421
QY 182 YGVN---NVKESDKSTLEEFARPOFLRMLEWYDDCYTRQKYLKDVOEKGS----N 234
DB 422 CKLVAVNIE-----POLYRMIREGRDYSELPLEYOKLKEKCDGKINTY 467
QY 235 DOLKCDI-ECKKCEDYVKTN-KKKKEW 260
DB 468 DKRYCVKPCOMACKSYDOMITRRKNOW 495

RESULT 11
US-10-153-273-17
Sequence 17, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun

Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:

LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-10-153-273-17

Query Match 14.2%; Score 218.5; DB 9; Length 324;
Best Local Similarity 24.7%; Pred. No. 1.4e-11;

Matches 60; Conservative 15; Mismatches 155; Indels 13; Gaps 2;

QY 28 CMPRRKCLVHFLANDNEIKKLOSQVNLKEAFISAAEFTFSYVYKSKDGEENELDK 87
DB 25 CIPPRQKCLHVLKXXXXXXKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 84
QY 88 ELKEGKIPPAFLRSMEYFGDYRDFLFGTDISKHGESEKLEQIDSLFKNGDOKSPNGK 147
DB 85 XXXXXD---FKROMFYTFADYRDLCTGDISSKKDTXXXXXXKXXXXXXXXXINSIR 140
QY 148 TROEWWTSHSEIMEAMICAL-----VKIGAKKDPFTENGVNVRKSDKSTLEE 198
DB 141 YKSMWETNGPYWGMICALKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 200
QY 199 FAKRPOFLRMLEWYDDCYTRQKYLKDVOEKCSNDQKCDTECNKCKEDYVAYMKKK 258
DB 201 XXXRPOFLRMLEWGENCKDQKKEYKVLAKCXXXXXXKXXXXXXCVACDKDCKOYH 260
QY 259 EMI 261
DB 261 SWI 263

RESULT 12
US-10-153-273-10
Sequence 10, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun

Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516

	Query Match	13.5%	Score 209;	DB 10;	Length 1501;
	Best Local Similarity	22.6%;	Pred. 6.1e-10;		
	Matches 65; Conservative	55;	Mismatches 115;	Indels 52;	Gaps 13
Oy	2 CHPKKNSNYPIWOC--GNINLYEDPRVCMPPRRQKLC---VHFLTANDNEIKKLISQAVML 56		:		:::: :

```

1      RESULT 14
2      US-10-153-273-15
3      : Sequence 15, Application US/10153273
4      : Patent No. US20020169305A1
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: Sim, Kim L.
9      :           Chitnis, Chetan
10     :           Miller, Louis H.
11     :           Peterson, David S.
12     :           Su, Xin-zhau
13     :           Wellens, Thomas E.
14     :
15     : TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
16     :                   AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
17     :
18     : NUMBER OF SEQUENCES: 37
19     :
20     : CORRESPONDENCE ADDRESS:
21     : ADDRESSEE: Knobbe Martens Olson & Bear
22     : STREET: 620 Newport Center Drive 16th Floor
23     : CITY: Newport Beach
24     : STATE: California
25     : COUNTRY: US
26     : ZIP: 92660
27     :
28     : COMPUTER READABLE FORM:
29     : MEDIUM TYPE: Floppy disk
30     : COMPUTER: IBM PC compatible
31     : OPERATING SYSTEM: PC-DOS/MS-DOS
32     : SOFTWARE: Patentin Release #1.0, Version #1.25.
33     :
34     : CURRENT APPLICATION DATA:
35     : APPLICATION NUMBER: US/10/153,273
36     : FILING DATE: 21-May-2002
37     : CLASSIFICATION: <Unknown>
38     :
39     : PRIOR APPLICATION DATA:
40     : APPLICATION NUMBER: US/09/210,288
41     : FILING DATE: <Unknown>
42     : ATTORNEY/AGENT INFORMATION:
43     : NAME: Fuller, Michael
44     : REGISTRATION NUMBER: 36,516
45     : REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
46     :
47     : TELECOMMUNICATION INFORMATION:
48     : TELEPHONE: (619) 235-8550
49     : TELEFAX: (619) 235-0176
50     :
51     : INFORMATION FOR SEQ ID NO: 15:
52     :
53     : SEQUENCE CHARACTERISTICS:
54     : LENGTH: 277 amino acids
55     : TYPE: amino acid
56     : STRANDEDNESS: single
57     : TOPOLOGY: linear
58     : MOLECULE TYPE: peptide
59     : HYPOTHETICAL: NO
60     : ANTI-SENSE: NO
61     : FRAGMENT TYPE: Internal
62     : ORIGINAL SOURCE:
63     : SEQUENCE DESCRIPTION: SEQ ID NO: 15:
64     :
65     : US-10-153-273-15

```


Result No.	Score	Query Match	Length	DB	ID	Description
1	276	100.0	3542	22	AA662142	P. falciparum FCRA
2	176	6.2	407	22	AA662148	P. falciparum varcru
3	112	4.3	351	22	AA662149	P. falciparum varcru
4	111	4.0	308	22	AA662151	P. falciparum varcru
5	111	4.0	311	22	AA662150	P. falciparum varcru
6	111	4.0	793	21	AA77911	Plasmodium DBL gen
7	111	4.0	934	16	AA70234	P. falciparum E31a
8	111	4.0	921	18	AAW22480	Plasmodium E31a
9	111	4.0	921	21	AAW77902	P. falciparum Ebl
10	10	3.6	32	22	AA662143	P. falciparum FCRA

121 GHGEGSLKEQIDSLFKNGDQKSPNGKTRQEWTEHSHEIWEAMLCALVKIGAKKDDFTE 180

Db 1399 GHEGSEKKEQIDSLFKNGDOKSPNGKTRQEWTEHSEIWEFAMCALVKGAKKDDFTE 1456
 QY 181 NGYNNVNFSDKSTLEEFAPKPOFLRWLTEMWDYCYTRQKYLKDVQEKCKSNDQKCD 240
 Db 1459 NGYNNVNFSDKSTLEEFAPKPOFLRWLTEMWDYCYTRQKYLKDVQEKCKSNDQKCD 1518
 QY 241 TECKNKCEDYVYMKRKKEMIPQDKYKDERDKRKF 276
 Db 1519 TECKNKCEDYVYMKRKKEMIPQDKYKDERDKRKF 1554

RESULT 2

AAB62148 ID AAB62148 standard; peptide: 407 AA.

AC AAB62148;

DT 29-MAY-2001 (first entry)

XX P. falciparum varCSA polypeptide A4 DBL4-gamma.

KW FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; A4 DBL4-gamma.

OS Plasmodium falciparum.

XX WO200116326-A2.

XX PD 08-MAR-2001.

XX PF 01-SEP-2000; 2000WO-US24195.

XX PR 01-SEP-1999; 99US-0152023.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 XX Gysin J, Pouvelle B, Fujii N, Smith J;
 XX WPI; 2001-235109/24.

XX DR WPI; 2001-235109/24.

XX PT Novel FCR3.varCSA protein, useful for modulating parasitized red blood
 XX cell binding, sequestration and onset of maternal malaria -

XX PS Disclosure; Page 72-73; 78pp; English.

XX CC The invention relates to a P. falciparum FCR3.varCSA protein, that is
 XX capable of binding to chondroitin sulfate A (CSA). The var gene and the
 XX corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
 XX modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 XX protein and the encoding gene are useful for treating and preventing
 XX CC maternal malaria in a patient afflicted with maternal malaria. The
 XX CC maternal malaria or in a patient afflicted with maternal malaria. The
 XX present sequence represents a P. falciparum varCSA polypeptide
 XX CC A4 DBL4-gamma.

XX SQ Sequence 407 AA;

Query Match 6.2%; Score 17; DB 22; Length 407;
 Best Local Similarity 100.0%; Pred. No. 1.6e-08;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GYRDFLFGTDISKGHG 123
 Db 128 GYRDFLFGTDISKGHG 144

RESULT 3

AAB62149 ID AAB62149 standard; peptide: 351 AA.
 XX

AC AAB62149;

XX DT 29-MAY-2001 (first entry)

XX P. falciparum varCSA polypeptide AttresDBL3-gamma.

KW FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; AttresDBL3-gamma.

OS Plasmodium falciparum.

XX WO200116326-A2.

XX PD 08-MAR-2001.

XX PF 01-SEP-2000; 2000WO-US24195.

XX PR 01-SEP-1999; 99US-0152023.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 XX Gysin J, Pouvelle B, Fujii N, Smith J;
 XX WPI; 2001-235109/24.

XX PT Novel FCR3.varCSA protein, useful for modulating parasitized red blood
 XX cell binding, sequestration and onset of maternal malaria -
 XX Claim 54; Page 73-74; 78pp; English.

XX CC The invention relates to a P. falciparum FCR3.varCSA protein, that is
 XX capable of binding to chondroitin sulfate A (CSA). The var gene and the
 XX corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
 XX modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 XX protein and the encoding gene are useful for treating and preventing
 XX CC maternal malaria in a patient afflicted with maternal malaria. The
 XX CC maternal malaria or in a patient afflicted with maternal malaria. The
 XX present sequence represents a P. falciparum varCSA polypeptide
 XX CC AttresDBL3-gamma.

XX SQ Sequence 351 AA;

Query Match 4.3%; Score 12; DB 22; Length 351;
 Best Local Similarity 100.0%; Pred. No. 0.0015;

Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 EAFIKSAAEETP 69
 Db 70 EAFIKSAAEETP 81

RESULT 4

AAB62151 ID AAB62151 standard; peptide: 308 AA.

XX AAB62151;

XX DT 29-MAY-2001 (first entry)

XX P. falciparum varCSA polypeptide ItG2-CS2 DBL2.

KW FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 KM erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; ItG2-CS2 DBL2..

OS Plasmodium falciparum.

XX WO200116326-A2.

XX PD 08-MAR-2001.

PF 01-SEP-2000; 2000MO-US24195.
 XX 01-SEP-1999; 9905-0152023.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujii N, Smith J;
 XX WPI; 2001-235109/24.
 DR Novel FCRA varCSA protein, useful for modulating parasitized red blood
 PT cell binding, sequestration and onset of maternal malaria -
 XX Claim 54; Page 75-76; 78pp; English.
 XX The invention relates to a P. falciparum FCRA varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents a P. falciparum varCSA polypeptide
 CC IIG2-CS2 DBL2.
 CC Sequence 308 AA;
 SQ
 Query Match 4.0%; Score 11; DB 22; Length 308;
 Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 28 CMPPRORLKV 38
 DB 34 CMPPRORLKV 44
 RESULT 5
 AAB62150
 ID AAB62150 standard; peptide; 311 AA.
 XX
 AC AAB62150;
 XX
 DT 29-MAY-2001 (first entry)
 XX
 DE P. falciparum varCSA polypeptide FCRA var3DBL-gamma.
 XX
 FCRA varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
 XX erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
 XX malaria; protozoacide; FCRA var3DBL-gamma.
 XX Plasmodium falciparum.
 XX WO200116326-A2.
 XX
 PD 08-MAR-2001.
 XX
 PF 01-SEP-2000; 2000MO-US24195.
 XX
 PR 01-SEP-1999; 9905-0152023.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
 PI Gysin J, Pouvelle B, Fujii N, Smith J;
 XX WPI; 2001-235109/24.
 DR Novel FCRA varCSA protein, useful for modulating parasitized red blood
 PT cell binding, sequestration and onset of maternal malaria -
 XX Disclosure; Page 74-75; 78pp; English.

CC The invention relates to a P. falciparum FCRA varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient afflicted at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents a P. falciparum varCSA polypeptide
 CC FCRA var3DBL-gamma.
 CC Sequence 311 AA;
 SQ
 Query Match 4.0%; Score 11; DB 22; Length 311;
 Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 28 CMPPRORLKV 38
 DB 36 CMPPRORLKV 46
 RESULT 6
 AAY7911
 ID AAY7911 standard; Protein; 324 AA.
 XX
 AC AAY7911;
 XX
 DT 13-JUN-2000 (first entry)
 XX
 DE Plasmodium DBL gene family E31a conserved domain.
 XX
 KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KW DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;
 XX protozoacide; E31a.
 XX Plasmodium sp.
 XX
 OS
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 1..324
 FT /note- "residues indicated Xaa are unspecified"
 XX
 PN US593827-A.
 XX
 PD 30-NOV-1999.
 XX
 PF 07-JUN-1995; 95US-0487826.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Slim KL, Chluis C, Peterson DS, Su X, Wellens TE, Miller LH;
 XX WPI; 2000-194198/17.
 DR Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX Examples; Fig 1; 93pp; English.
 XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 CC Binding Protein (SAMP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SAMP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria.

SQ Sequence 324 AA;

Query Match 4.0%; Score 11; DB 21; Length 324;
 Best Local Similarity 100.0%; Pred. No. 0.014;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLTEM 212
 |||||
 DB 204 RPOFLRWLTEM 214

RESULT 7

ID AAR70234 standard; Protein: 793 AA.

AC AAR70234;

DT 22-SEP-1995 (first entry)

DE P. falciparum E31a.

Erythrocyte binding ligand; E31a; binding domain; malaria; therapy;
 vaccine.

XX Plasmodium falciparum.

XX WO9507353-A.

XX 16-MAR-1995.

XX 07-SEP-1994; 94WO-US10230.

XX 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

XX Wellens TE;

XX WPI: 1995-123427/16.

XX N-PSDB; AA083527.

XX New erythrocyte binding domain polypeptide(s) - isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and
 PT prevention of malaria

XX Disclosure; Page 51-52; 81pp; English.

XX Erythrocyte binding ligand (EBL) family genes were cloned from
 P. falciparum chromosome 7 subsegment libraries constructed during
 genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AA083526), E31a (AA083527), EBL-e2 (AA083528) and ProJ3
 CC (AA083529), encode the proteins given in AAR70233-36, respectively.
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.

SQ Sequence 793 AA;

Query Match 4.0%; Score 11; DB 16; Length 793;
 Best Local Similarity 100.0%; Pred. No. 0.031;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLTEM 212
 |||||
 DB 618 RPOFLRWLTEM 628

RESULT 8

ID AAM22480 standard; Protein: 921 AA.

AC AAM22480;

XX 07-OCT-1997 (first entry)
 DT XX
 DE Plasmodium E31a.

XX DBL gene family; SABP; stalic acid binding protein; vaccine; therapy;
 XX Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.

XX Plasmodium falciparum.

XX WO9640766-A2.

XX 19-DEC-1996.

XX 07-JUN-1996; 96WO-US09508.

XX 07-JUN-1995; 95US-0487826.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

XX Wellens TE;

XX WPI: 1997-052231/05.

XX N-PSDB; AAT72895.

XX New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stalic acid
 PT binding proteins

XX Disclosure; Page 43-45; 96pp; English.

XX This sequence represents E31a of Plasmodium. E31a belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stalic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).

SQ Sequence 921 AA;

Query Match 4.0%; Score 11; DB 18; Length 921;
 Best Local Similarity 100.0%; Pred. No. 0.036;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLTEM 212
 |||||
 DB 618 RPOFLRWLTEM 628

RESULT 9

ID AAT77902 standard; Protein: 921 AA.

AC AAT77902;

DT 13-JUN-2000 (first entry)

```
DE P. falciparum ebl-1 related polypeptide.
XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
KW DABP; Stalic Acid Binding Protein; SABB; malaria; vaccine; Immunisation;
RV protozoacide.
XX
XX OS Plasmodium falciparum.
XX US9593827-A.
XX
XX 30-NOV-1999.
XX
XX 07-JUN-1995; .95US-0487826.
XX
XX 10-SEP-1993; 93US-0119677.
XX
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Slim XL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
R1 WPI; 2000-194198/17.
DR N-PSDB; AAZ98284.
XX
XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
PT falciparum erythrocyte binding proteins useful for vaccinating against
PT malaria -
XX
XX Disclosure; Columns 61-66; 93bp; English.
XX
XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Stalic Acid
CC Binding Protein (SABB), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABB are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and stalic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria.
XX
XX Sequence 921 AA;
SQ
Query Match 4.0%; Score 11; DB 21; Length 921;
Best Local Similarity 100.0%; Pred. No. 0.036;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
202 RPOFLRWLTW 212
|||||||
618 RPOFLRWLTW 628
RESULT 10
AAB62143
ID AAB62143 standard; peptide: 32 AA.
XX
XX AAB62143;
AC
XX
XX 29-MAY-2001 (first entry)
DT
XX
XX P. falciparum FCR3.varCSA protein fragment.
DE
XX
XX FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene; PFEMP1;
KW erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
KW malaria; protozoacide.
XX
XX Plasmodium falciparum.
OS
XX
XX WO200116326-A2.
PN
XX
XX 08-MAR-2001.
PD
XX
XX 01-SEP-2000; 2000WO-US24195.
PF
```

XX	01-SEP-1999;	99US-0152023.
FR	(USSH) US DEPT HEALTH & HUMAN SERVICES.	
XX		
PA	Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;	
PI	Gysin J, Pouvelle B, Fujii N, Smith J;	
XX		
DR	WPI; 2001-235109/24.	
XX		
PT	Novel FCRI3.varCSA protein, useful for modulating parasitized red blood	
PT	cell binding, sequestration and onset of maternal malaria -	
XX		
PS	Claim 25; Page 21; 78pp: English.	
XX		
CC	The invention relates to a P. falciparum FCRI3.varCSA protein, that is	
CC	capable of binding to chondroitin sulfate A (CSA). The var gene and the	
CC	corresponding P.falciparum erythrocyte membrane protein 1 (PFEMP1)	
CC	modulate adhesion of parasitized red blood cell (PRBC) to CSA. The	
CC	protein and the encoding gene are useful for treating and preventing	
CC	maternal malaria in a patient identified at a risk for contracting	
CC	maternal malaria or in a patient afflicted with maternal malaria. The	
CC	present sequence represents a fragment of the P. falciparum FCRI3.varCSA	
XX	protein.	
SO	Sequence 32 AA;	
Query Match	3.6%; Score 10; DB 22; Length 32;	
Best Local Similarity	100.0%; Pred. No. 0.019;	
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		
OY 102 MFYTFGDYRD 111		
Db 19 MFYTFGDYRD 28		
RESULT 11		
AA77915		
ID AA77915 standard; Protein; 311 AA.		
XX		
AC AA77915;		
XX		
DT 13-JUN-2000 (first entry)		
XX		
DE Plasmodium DBL gene family Pro3 conserved domain F3.		
XX		
DDBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;		
KW DABP; Sialic Acid Binding Protein; SAMP; malaria; vaccine; immunisation;		
KW protozoacide; Pro3.		
XX		
OS Plasmodium sp.		
XX		
FH Key Location/Qualifiers		
FT Misc-difference 1..311		
FT /note- "residues indicated Xaa are unspecified"		
XX		
PN US593827-A.		
XX		
PD 30-NOV-1999.		
XX		
PF 07-JUN-1995; 95US-0487826.		
XX		
PR 10-SEP-1993; 93US-0119677.		
XX		
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.		
XX		
PI Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;		
XX		
DR WPI; 2000-194198/17.		
XX		
PT Isolated protein binding domains from Plasmodium vivax and Plasmodium		
PT falciparum erythrocyte binding proteins useful for vaccinating against		
FT malaria -		

XX PS Examples; Fig 1; 93pp; English.
XX CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
CC identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
CC Binding Protein (SABP), which are soluble proteins that appear in the
CC culture supernatant after erythrocytes infected with malaria release
CC merozoites. Immunochemical studies indicate that DABP and SABP are the
CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
CC and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
CC used to vaccinate against malaria, especially caused by P. falciparum.
CC Immunization with the polypeptide provides effective protection against
CC malaria.
XX SQ Sequence 311 AA;
Query Match 3.6%; Score 10; DB 21; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
28 CMPPRKQKLC 37
|||||
24 CMPPRKQKLC 33
RESULT 12
AAR70236
ID AAR70236 standard; Protein; 2703 AA.
XX AC AAR70236;
XX DT 22-SEP-1995 (first entry)
XX DE P. falciparum Proj3.
XX KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
XX KM vaccine.
XX OS Plasmodium falciparum.
XX PN WO9507353-A.
XX PD 16-MAR-1995.
XX PF 07-SEP-1994; 94WO-US10230.
XX PR 10-SEP-1993; 93US-0119677.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
DR WPI; 1995-123427/16.
DR N-PSDB; AAQ83529.
XX PI New erythrocyte binding domain polypeptide(s) - isolated from
PI Plasmodium binding proteins, used in diagnosis, treatment and
PI prevention of malaria
XX PS Disclosure; Page 61-65; 81pp; English.
XX CC Erythrocyte binding ligand (EBL) family genes were cloned from
CC P. falciparum chromosome 7 subsegment libraries constructed during
CC genetic studies of the chloroquine resistance locus. The 4 genes,
CC EBL-e1 (AAQ83526), E31a (AAQ83527), EBL-e2 (AAQ83528) and Proj3
CC (AAQ83529), encode the proteins given in AAR70233-36, respectively. The
CC binding domains of such proteins can be expressed e.g. in E. coli,
CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
CC cells, and provide protection against P. falciparum.
SQ Sequence 2703 AA;

Query Match 3.6%; Score 10; DB 16; Length 2703;
Best Local Similarity 100.0%; Pred. No. 0.93;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
28 CMPPRKQKLC 37
|||||
1384 CMPPRKQKLC 1393
Db 1384 CMPPRKQKLC 1393
RESULT 13
AAW22482
ID AAW22482 standard; Protein; 2710 AA.
XX AC AAW22482;
XX DT 07-OCT-1997 (first entry)
XX DE Plasmodium Proj3.
XX KW DBL gene family; SABP; stialic acid binding protein; vaccine; therapy;
XX KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
XX KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
XX KM Plasmodium.
XX OS Plasmodium falciparum.
XX PN WO9640766-A2.
XX PD 19-DEC-1996.
XX PF 07-JUN-1996; 96WO-US09508.
XX PR 07-JUN-1995; 95US-0487826.
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
PI Wellens TE;
DR WPI; 1997-052231/05.
DR N-PSDB; AAT72897.
XX PI New malaria vaccines - contains cysteine-rich DBL family protein
PI binding domains homologous domains of the Duffy and stialic acid
PI binding proteins
XX PS Disclosure; Page 50-56; 96pp; English.
XX CC This sequence represents Proj3 of Plasmodium. Proj3 belongs to
CC the Duffy binding like (DBL) family of genes which have homology to the
CC Duffy antigen binding protein (DABP) and stialic acid binding protein
CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
CC var family of genes modulate cytoadherence and antigenic variation of
CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
CC protein (DABP) are soluble proteins that appear in the culture
CC supernatant after infected erythrocytes release merozoites. DABP and
CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
CC surface. These proteins are necessary for erythrocyte invasion by the
CC parasite. This sequence can be used in the compositions of the
CC invention. The compositions are for the treatment and prevention of
CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family. A
CC family of genes having homology with conserved regions of DABP and SABP.
CC The compositions are used for the treatment and prevention of malaria.
CC They are also used in the preparation of vaccines for inducing a
CC protective immune response in a mammal to Plasmodium merozoites
CC (especially Plasmodium falciparum or Plasmodium vivax).
SQ Sequence 2710 AA;
Query Match 3.6%; Score 10; DB 18; Length 2710;
Best Local Similarity 100.0%; Pred. No. 0.94;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 28 CMPPRRQKLC 37
 |||||
 Db 1391 CMPPRRQKLC 1400

RESULT 14
 AAY77904
 ID AAY77904 standard; Protein; 2710 AA.

XX AAY77904;

DT 13-JUN-2000 (first entry)

DE P. falciparum Proj3 binding domain polypeptide.

KM DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KM DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 KM Protozoa; Proj3.

XX Plasmodium falciparum.

XX US5993827-A.

XX 30-NOV-1999.

XX 07-JUN-1995; 95US-0487826.

XX 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;

XX WPI; 2000-194198/17.

XX DR N-PSDB; AA298286.

XX Isolated protein binding domains from Plasmodium vivax and Plasmodium

XX falciparum erythrocyte binding proteins useful for vaccinating against

XX malaria -

XX Disclosure; Columns 79-92; 93pp; English.

XX The invention relates to ebl-1 polypeptides that are encoded by the DBL

XX (Duffy-binding like) gene family. The ebl-1 proteins are substantially

XX identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid

XX Binding Protein (SABP), which are soluble proteins that appear in the

XX culture supernatant after erythrocytes infected with malaria release

XX merozoites. Immunochemical studies indicate that DABP and SABP are the

XX respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy

XX CC and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be

XX CC used to vaccinate against malaria, especially caused by P. falciparum.

XX CC Immunization with the polypeptide provides effective protection against

XX CC malaria. The present sequence represents the Proj3 binding domain

XX CC polypeptide.

XX SQ Sequence 2710 AA;

XX Query Match 3.6%; Score 10; DB 21; Length 2710;

XX Best Local Similarity 100.0%; Pred. No. 0.94;

XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 12-SEP-1997 (first entry)
 DT Plasmodium var-7.
 XX
 DE Plasmodium var-7.
 XX

KM DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KM Plasmodium.

XX Plasmodium vivax.

XX OS Plasmodium falciparum.

XX WO9640766-A2.

XX 19-DEC-1996.

XX 07-JUN-1996; 96WO-US09508.

XX 07-JUN-1995; 95US-0487826.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;

XX WPI; 1997-052231/05.

XX DR N-PSDB; AAT72882.

XX New malaria vaccines - contains cysteine-rich DBL family protein

XX binding domains homologous domains of the Duffy and sialic acid

XX binding proteins

XX Claim 8; Page 61-67; 96pp; English.

XX This sequence represents var-7 of Plasmodium. Var-7 belongs to

XX the Duffy binding like (DBL) family of genes which have homology to the

XX Duffy antigen binding protein (DABP) and sialic acid binding protein

XX (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The

XX var family of genes modulate cytoadherence and antigenic variation of

XX Plasmodium infected erythrocytes. SABP and the Duffy antigen binding

XX protein (DABP) are soluble proteins that appear in the culture

XX supernatant after infected erythrocytes release merozoites. DABP and

XX SABP mediate the binding of merozoites and schizonts to the erythrocyte

XX surface. These proteins are necessary for erythrocyte invasion by the

XX parasite. This sequence can be used in the compositions of the

XX invention. The compositions are for the treatment and prevention of

XX malaria, and comprise either a nucleotide sequence or encoded polypeptide

XX of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a

XX family of genes having homology with conserved regions of DABP and SABP.

XX The compositions are used for the treatment and prevention of malaria.

XX They are also used in the preparation of vaccines for inducing a

XX protective immune response in a mammal to Plasmodium merozoites

XX (especially Plasmodium falciparum or Plasmodium vivax).

XX SQ Sequence 3060 AA;

XX Query Match 3.6%; Score 10; DB 18; Length 3060;

XX Best Local Similarity 100.0%; Pred. No. 1;

XX Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX QY 28 CMPPRRQKLC 37
 |||||
 Db 1389 CMPPRRQKLC 1398

Search completed: April 28, 2003, 10:49:46
 Job time : 51.6111 secs

100

100

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:48:21 ; Search time 24.8889 Seconds

(without alignments)
1066.062 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPKKNSNGYDWCQGNIN.....KKEMIPDKYRDERDKRPF 276

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 263224 seqs, 96134422 residues

rd size : 9

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	17	6.2	3026	2 T28431	variant surface pr
2	11	4.0	3006	2 T28625	variant-specific s
3	10	3.6	1711	2 C71625	variant-specific s
4	10	3.6	3078	2 T28432	variant-specific s

ALIGNMENTS

RESULT 1

T28431 variant surface protein 1 homolog A4VAR - malaria parasite (Plasmodium falciparum) (frag

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28431

R:Smith, J.D.; Kyes, S.; Craig, A.G.; Fegan, T.; Hudson-Taylor, D.; Miller, L.H.; Barruch
Mol. Biochem. Parasitol. 97, 133-148, 1998

A:Title: Analysis of adhesive domains from the A4VAR Plasmodium falciparum erythrocyte m

A:Reference number: 220486; MUID:99094502; PMID:9879893

A:Accession: T28431

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3026 <SMT>

A:Cross-references: EMBL:LA2244; NID:93540144; PID:93540145; PIDN:AA03351.1

C:Genetics:
A:Gene: var

Query Match

6.2%; Score 17; DB 2; Length 3026;

Best Local Similarity 100.0%; Pred. No. 1.3e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GDYRDFLEGTDISKGHG 123
|||||

Db 2122 GDYRDFLEGTDISKGHG 2138

RESULT 2

T28625

variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28625

R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfelde, J.A.; Peterson,
Cell 82, 89-100, 1995

A:Title: The large diverse gene family var encodes proteins involved in cytoadherence

A:Reference number: 220487; MUID:95330813; PMID:7606788

A:Accession: T28625

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3006 <SU>

A:Cross-references: EMBL:LA0609; NID:9886376; PID:9886377; PIDN:AA75397.1

C:Genetics:
A:Gene: var-3

A:Introns: 2597/3

Query Match
Best Local Similarity 100.0%; Pred. No. 0.022;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CMPPRKRLCV 38
|||||

Db 1270 CMPPRKRLCV 1280

RESULT 3

C71625

variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium fa

C:Species: Plasmodium falciparum

C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000

C:Accession: C71625

R:Gerder, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.
J.; Perera, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H

Science 287, 1126-1132, 1998

A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.

A:Reference number: A71600; MUID:99021743; PMID:9804551

A:Accession: C71625

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1711 <GAR>

A:Cross-references: GB:AE001366; GB:AE001362; NID:93845070; PIDN:AA071792.1; PID:9384

A:Experimental source: clone 3D7

C:Genetics:
A:Gene: PFB0010w

Query Match
Best Local Similarity 100.0%; Pred. No. 0.15;

Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 MFYFGDYRD 111
|||||

Db 951 MFYFGDYRD 960

RESULT 4

T28432

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)

N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)

C:Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000

C:Accession: T28432

R:Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herfelde, J.A.; Peterson,

Cell 82, 89-100, 1995
 A>Title: The large diverse gene family var encodes proteins involved in cytoadherence at
 A:Reference number: Z20487; MID:95330813; PMID:7606788
 A:Accession: T28432
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-3078 <STX>
 A:Cross-references: EMBL:L40608; NID:9886374; PID:9886375; PIDN:AAA75396.1
 C:Gene: var-1
 A:Gene: var-1
 A:introns: 2611/3

Query Match 3.6%; Score 10; DB 2; Length 3078;
 Best Local Similarity 100.0%; Pred. No. 0.25;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 28 CMPPRRQKLC 37
 |||||
 Db 1389 CMPPRRQKLC 1398

Search completed: April 28, 2003, 10:52:29
 Search time: 26.8889 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:35:40 ; Search time 14.2222 Seconds
(without alignments)
804.900 Million cell updates/sec

Title: us-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPRKNSNGYDPDQGNIN.....KKEMIPQDKRYKDERDKRRF 276

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues
Hit size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database : SwissProt_40:*

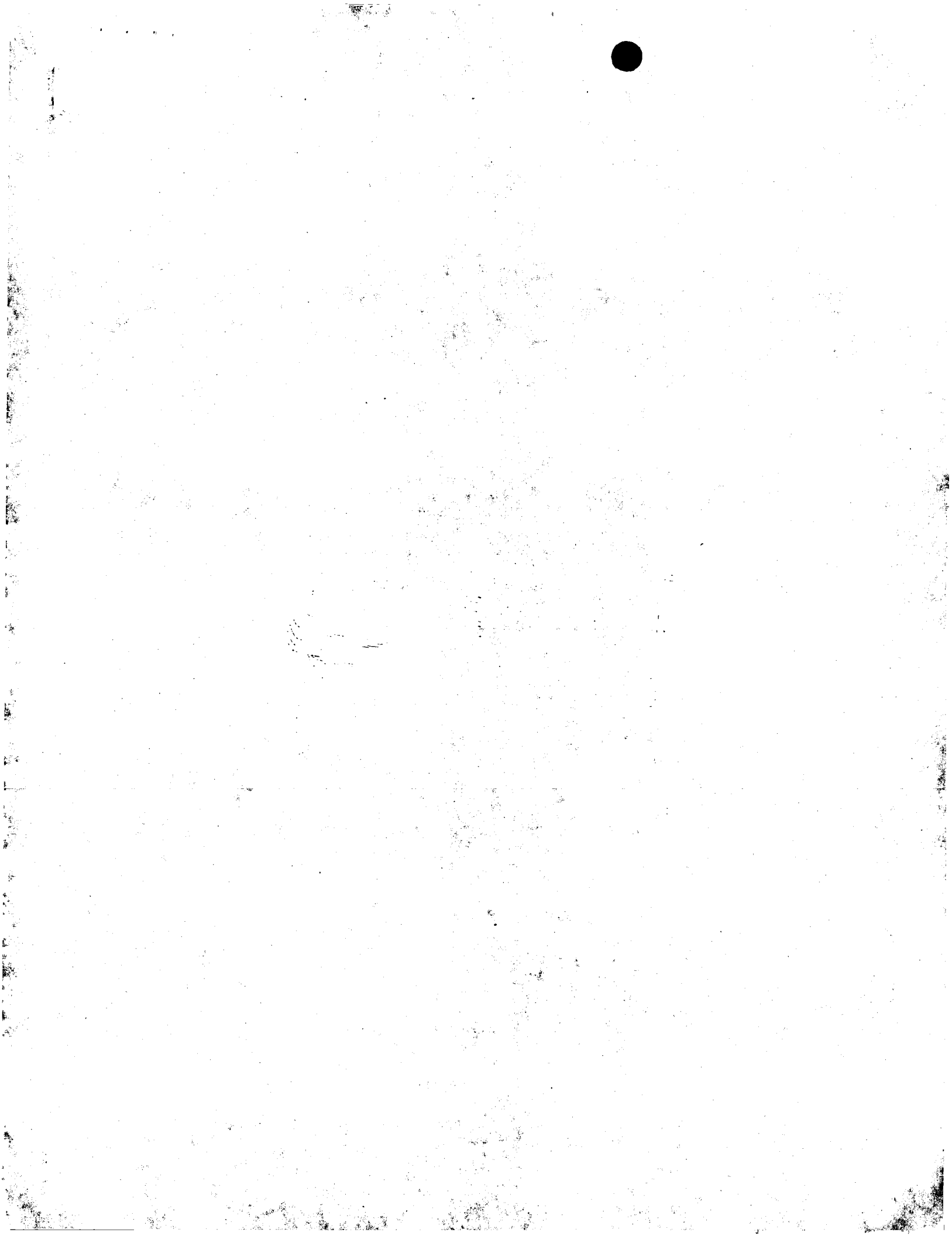
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description

No matches found					

Search completed: April 28, 2003, 10:50:21
Job time : 14.2222 secs



GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:45:26 ; Search time 25.333 Seconds

(without alignments)
2244.829 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPKKNSNGYPMQCGNIN.....KKEMIPQDKYKDERDKRRF 276

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

rd size : 9

Total number of hits satisfying chosen parameters: 19

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

- 1: SP-archaea:*
- 2: SP-bacteria:*
- 3: SP-fungi:*
- 4: SP-human:*
- 5: SP-invertebrate:*
- 6: SP-mammal:*
- 7: SP-mhc:*
- 8: SP-organelle:*
- 10: SP-phage:*
- 11: SP-plant:*
- 12: SP-rodent:*
- 13: SP-virus:*
- 14: SP-vertebrate:*
- 15: SP-unclassified:*
- 16: SP-virus:*
- 17: SP-bacteriophage:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	276	100.0	3542	09U5M2	09U5M2 plasmodium
2	20	7.2	1615	08T325	08T325 plasmodium
3	20	7.2	3287	08T326	08T326 plasmodium
4	17	6.2	169	025990	025990 plasmodium
5	17	6.2	182	000832	000832 plasmodium
6	17	6.2	3026	026030	026030 plasmodium
7	14	5.1	173	025988	025988 plasmodium
8	12	4.3	1685	09U4A2	09U4A2 plasmodium
9	11	4.0	162	09B1E7	09B1E7 plasmodium
10	11	4.0	195	000831	000831 plasmodium
11	11	4.0	921	025989	025989 plasmodium
12	11	4.0	2527	095W83	095W83 plasmodium
13	11	4.0	2710	09XZB8	09XZB8 plasmodium
14	11	4.0	3006	026032	026032 plasmodium
15	10	3.6	81	09NFB4	09NFB4 plasmodium
16	10	3.6	1711	096108	096108 plasmodium

17	10	3.6	2658	5	08T5G0	08T5G0 plasmodium
18	10	3.6	3078	5	Q26031	Q26031 plasmodium
19	9	3.3	1327	5	Q9NFB4	Q9NFB4 plasmodium

ALIGNMENTS

RESULT 1

ID	Query Match	Score	DB 5	Length	3542
09U5M2	PRELIMINARY:	PRT:	3542	AA.	
AC	09U5M2:				
DT	01-MAY-2000 (TREMblrel. 13, Created)				
DT	01-MAY-2000 (TREMblrel. 13, Last sequence update)				
DT	01-MAR-2002 (TREMblrel. 20, Last annotation update)				
DE	FCR3 CSA ligand (Fragment).				
GN	VAR.				
OS	Plasmodium falciparum.				
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.				
OX	NCBI_TaxID=5633;				
RN	SEQUENCE FROM N.A.				
RP	STRAIN=FCR3;				
RC	MEDLINE=20006305; PubMed=10535993;				
RA	Buffet P., Gamain B., Scheldig C., Baruch B., Oishi S., Fujii N.,				
RA	Fusai T., Parzy D., Miller L.H., Gysin J., Schert A.;				
RT	"Plasmodium falciparum domain mediating adhesion to Chondroitin				
RL	sulfate A: A receptor for human placental infection."				
DR	Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).				
DR	EMBL: AJ133811; CAB59840.1;				
DR	InterPro: IPR001219; Neurotoxin.				
DR	InterPro: IPR004258; PEEMP.				
DR	Pfam: PF03011; PEEMP; 1.				
DR	PRINTS: PRO0284; TOXIN.				
FT	NON TER. 3542 3542				
SO	SEQUENCE 3542 AA; 413089 MW; 970D85E88DA2EC2 CRC64;				

Query Match 100.0%; Score 276; DB 5; Length 3542;
Best Local Similarity 100.0%; Pred. No. 2.9e-284;
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	DCHPKKNSNGYPMQCGNINLYEDPRVCMPPRRQKLCVHFLANDNEIKKLSQVNLKNAF	60
DB	1279	DCHPKKNSNGYPMQCGNINLYEDPRVCMPPRRQKLCVHFLANDNEIKKLSQVNLKNAF	1338
QY	61	IKSAAAEFFFSWYKSKDGEGNELDKELKEKIPAFIRSMFYFGDYRDLFTGTDISK	120
DB	1339	IKSAAAEFFFSWYKSKDGEGNELDKELKEKIPAFIRSMFYFGDYRDLFTGTDISK	1398
QY	121	GHGEGSKLEQIDSLFKNGDQSPNGKTRQEMWTEHSHEIWEAMICALVKGAKKDPETE	180
DB	1399	GHGEGSKLEQIDSLFKNGDQSPNGKTRQEMWTEHSHEIWEAMICALVKGAKKDPETE	1458
QY	181	NYGVNNVKSFSKSTLEEFARPOFLRLTEYDYCYTRQKYLTDVQEKCSNOIKCD	240
DB	1459	NYGVNNVKSFSKSTLEEFARPOFLRLTEYDYCYTRQKYLTDVQEKCSNOIKCD	1518
QY	241	TECNKCCEDYKYMKKKEMIPQDKYKDERDKRRF	276
DB	1519	TECNKCCEDYKYMKKKEMIPQDKYKDERDKRRF	1554
RESULT 2			
ID	08T325	PRELIMINARY:	PRT: 1615
AC	08T325:		AA.
DT	01-JUN-2002 (TREMblrel. 21, Created)		
DT	01-JUN-2002 (TREMblrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMblrel. 21, Last annotation update)		
DE	pFEMP1 (Fragment).		
GN	TM284VAR3.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		

OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM284;
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implied in malaria in pregnancy.";
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420412; CAD20868.1; -
FT NON_TER 1615 1615
SQ SEQUENCE 1615 AA; 188639 MW; FCA896C00B86DEA6 CRC64;

Query Match 7.2%; Score 20; DB 5; Length 1615;
Best Local Similarity 100.0%; Pred. No. 5.2e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 231 CKSNDQKCDTECNKCCDY 250
Db 1495 CKSNDQKCDTECNKCCDY 1514
|||||
SOUT 3
ID 08T326 PRELIMINARY; PRT; 3287 AA.
AC 08T326;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE PLEMP1 (Fragment).
GN TM180VAR2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TM180;
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implied in malaria in pregnancy.";
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420411; CAD20867.1; -
FT NON_TER 3287 3287
SQ SEQUENCE 3287 AA; 383550 MW; 58F8F866FC244536 CRC64;

Query Match 7.2%; Score 20; DB 5; Length 3287;
Best Local Similarity 100.0%; Pred. No. 9.7e-12;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

231 CKSNDQKCDTECNKCCDY 250
Db 1495 CKSNDQKCDTECNKCCDY 1514
|||||
RESULT 4
ID 025990 PRELIMINARY; PRT; 169 AA.
AC 025990;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-NOV-1996 (TREMBlrel. 08, Last annotation update)
DE (clone pNM4) ORF (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2/NM;
RX MEDLINE=95350219; PubMed=7624377;
RA Peterson D.S., Miller I.H., Wellens T.E.;
RT "Isolation of multiple sequences from the Plasmodium falciparum genome
that encode conserved domains homologous to those in erythrocyte-

RT binding proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
DR EMBL; L38455; AAC37241.1; -
FT NON_TER 1 1
FT NON_TER 169 169
SQ SEQUENCE 169 AA; 19120 MW; BCC6B0FC3E1380 CRC64;

Query Match 6.2%; Score 17; DB 5; Length 169;
Best Local Similarity 100.0%; Pred. No. 1.1e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GDYRDFLEFGTIDSKGHG 123
Db 71 GDYRDFLEFGTIDSKGHG 87
|||||
RESULT 5
ID 000832 PRELIMINARY; PRT; 182 AA.
AC 000832;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Partial erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum (Isolate Dd2).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=57267;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DD2;
RX MEDLINE=97342604; PubMed=9199301;
RA Fischer K., Horrocks P., Preuss M., Wiesner J., Wensch S.,
RA Camargo A.A., Lanzer M.;
RT "Expression of var genes located within polymorphic subtelomeric
domains of Plasmodium falciparum chromosomes.";
RL Mol. Cell. Biol. 17:3679-3686(1997).
DR EMBL; Y11910; CAA72661.1; -
FT NON_TER 1 1
FT NON_TER 182 182
SQ SEQUENCE 182 AA; 20830 MW; 197CA46778AD148B CRC64;

Query Match 6.2%; Score 17; DB 5; Length 182;
Best Local Similarity 100.0%; Pred. No. 1.2e-09;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GDYRDFLEFGTIDSKGHG 123
Db 78 GDYRDFLEFGTIDSKGHG 94
|||||
RESULT 6
ID 026030 PRELIMINARY; PRT; 3026 AA.
AC 026030;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Variant surface protein (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN=IT 4/25/5;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfeldt J.A.,
RA Peterson D.S., Ravetch J.A., Wellens T.E.;
RT "The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium falciparum-
infected erythrocytes.";
RL Cell 82:89-100(1995).

[2]
RN SEQUENCE OF 2044-2922 FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.D., Chitnis C.E., Craig A.G., Roberts D.J.,
RA Hudson-Taylor D.E., Peterson D.S., Pinches R., Newbold C.I.,
RA Miller L.H.;
RT "Switches in the expression of Plasmodium falciparum var genes
correlate with changes in antigenic and cytoadherent phenotypes of
infected erythrocytes.";
RT Cell 0:0-0(1995).
RN [3]
RC SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA MDLINE-99094502;
RA Smith J.D., Kyes S., Craig A.G., Fagan T., Hudson-Taylor D.,
RA Miller L.H., Baruch D.I., Newbold C.I.;
RT "Analysis of adhesive domains from the AVAR Plasmodium falciparum
erythrocyte membrane protein-1 identifies a CD36 binding domain.";
RT Mol. Biochem. Parasitol. 97:133-148(1998).
RN [4]
RC SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Smith J.;
RN Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
RN [5]
RC SEQUENCE FROM N.A.
RC STRAIN-IT 4/25/5;
RA Kyes S., Smith J.;
RN Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: LA2244; AAD03351.1; -;
DR InterPro: IPR004236; PFEMP.
DR Pfam: PF03011; PFEMP; 2.
FT NON_TER 3026 3026
SQ SEQUENCE 3026 AA; 348124 MW; EC0A971A82193788 CRC64;

Query Match 6.28; Score 17; DB 5; Length 3026;
Best Local Similarity 100.0%; Pred. No. 1.4e-08;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 GDYRDLFGTDISKHG 123
Db 2122 GDYRDLFGTDISKHG 2138
|||||

RESULT 7
Q25988 PRELIMINARY; PRT; 173 AA.
ID Q25988;
AC 025988;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)
DE (clone pM5) ORF (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DD2/NM;
RX MEDLINE-95350219; Pubmed-7624377;
RA Peterson D.S., Miller L.H., Wellens T.E.;
RT "Isolation of multiple sequences from the Plasmodium falciparum genome
that encode conserved domains homologous to those in erythrocyte-
binding proteins.";
RT Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
DR EMBL: L38453; AAC37239.1; -;
FT NON_TER 1 1
FT NON_TER 173 173
SQ SEQUENCE 173 AA; 19828 MW; 365051FBF25ED2F2 CRC64;

Query Match 5.18; Score 14; DB 5; Length 173;
Best Local Similarity 100.0%; Pred. No. 1.8e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 RDLFGTDISKHG 123
Db 70 RDLFGTDISKHG 83
|||||

RESULT 8
ID Q904A2 PRELIMINARY; PRT; 1685 AA.
AC Q904A2;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Variant surface protein PfEMP1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-IT;
RX MEDLINE-20144115; Pubmed-10677532;
RA Smith J.D., Craig A.G., Kriek N., Hudson-Taylor D., Kyes S., Fagan T.,
RA Pinches R., Baruch D.I., Newbold C.I., Miller L.H.;
RT "Identification of a Plasmodium falciparum intercellular adhesion
molecule-1 binding domain: A parasite adhesion trait implicated in
cerebral malaria.";
RT Proc. Natl. Acad. Sci. U.S.A. 97:1766-1771(2000).
DR EMBL: AF193424; AAF18980.1; -;
DR InterPro: IPR000454; Eub_ATPase_Csub.
DR InterPro: IPR004236; PFEMP.
DR Pfam: PF03011; PFEMP; 1.
DR PROSITE: PS00605; ATPASE_C; 1.
FT NON_TER 1685 1685
SQ SEQUENCE 1685 AA; 193671 MW; D1FD42666B0551E CRC64;

Query Match 4.38; Score 12; DB 5; Length 1685;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 58 EAFKSAARET 69
Db 1293 EAFKSAARET 1304
|||||

RESULT 9
Q9BJE7 PRELIMINARY; PRT; 162 AA.
ID Q9BJE7;
AC Q9BJE7;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-720;
RX MEDLINE-21136462; Pubmed-11237850;
RA Khattab A., Kun J., Deloron P., Kremsner P.G., Klinkert M.O.;
RT "Variants of Plasmodium falciparum erythrocyte Membrane Protein 1
Expressed by Different Placental Parasites are Closely Related and
Adhere to Chondroitin Sulfate A.";
RT J. Infect. Dis. 183:1165-1169(2001).
DR EMBL: AF334806; AAK28129.1; -;
FT NON_TER 1 1
FT NON_TER 162 162
SQ SEQUENCE 162 AA; 18788 MW; BC2E24182BF1078E CRC64;

Query Match 4.08; Score 11; DB 5; Length 162;
Best Local Similarity 100.0%; Pred. No. 0.0026;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 AFKSAAEFT 69
 DB 25 AFKSAAEFT 35

RESULT 10

000831 PRELIMINARY; PRT; 195 AA.
 AC 000831;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Partial erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum (isolate Dd2).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=57267;
 RN [1]
 RC STRAIN-DD2;
 MEDLINE-97342604; PubMed-9199301;
 Fischer K., Horrocks P., Preuss M., Wiesner J., Wuenesch S.,
 Camargo A.A., Lanzer M.;
 "Expression of var genes located within polymorphic subtelomeric
 domains of Plasmodium falciparum chromosomes.";
 RT Mol. Cell. Biol. 17:3679-3686(1997).
 DR EMBL; Y11909; CAA72660.1; -.
 FT NON_TER 1
 FT 1
 SQ SEQUENCE 195 AA; 22667 MW; 8B5F1D5F18366F21 CRC64;

Query Match
 Best Local Similarity 4.0%; Score 11; DB 5; Length 195;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 110 RDEFGTDISK 120
 DB 88 RDEFGTDISK 98

RESULT 11

025989 PRELIMINARY; PRT; 921 AA.
 AC 025989;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Plasmodium falciparum.
 (clone pS31H) ORF (Fragment).
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5833;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-DD2/MM;
 MEDLINE-95350219; PubMed-7624377;
 RA Peterson D.S., Miller L.H., Wellens T.E.;
 "Isolation of multiple sequences from the Plasmodium falciparum genome
 binding proteins.";
 RT that encode conserved domains homologous to those in erythrocyte-
 RT binding proteins.";
 RU Proc. Natl. Acad. Sci. U.S.A. 92:7100-7104(1995).
 DR EMBL; J38454; AAC37240.1; -.
 DR InterPro; IPR004258; PFEMP.
 DR Pfam; PF03011; PFEMP. 1.
 FT NON_TER 1
 FT 1
 SQ SEQUENCE 921 AA; 104711 MW; D98BC07763628746 CRC64;

Query Match
 Best Local Similarity 4.0%; Score 11; DB 5; Length 921;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLTEM 212
 DB 618 RPOFLRWLTEM 628

RESULT 12

095W83 PRELIMINARY; PRT; 2527 AA.
 AC 095W83;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR1.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RC SEQUENCE FROM N.A.
 RX MEDLINE-21442075; PubMed-11557894;
 RA Flick K., Scholander C., Chen Q., Fernandez V., Pouvelle B., Gysin J.,
 Wahlgren M.;
 "Role of nonimmune IgG bound to PfEMP1 in placental malaria.";
 RT Science 293:2098-2100(2001).
 DR EMBL; AF366567; AAL12845.1; -.
 DR InterPro; IPR004258; PFEMP.
 DR Pfam; PF03011; PFEMP. 2.
 FT NON_TER 2527
 FT 2527
 SQ SEQUENCE 2527 AA; 292137 MW; 59EE2766BF0425E4 CRC64;

Query Match
 Best Local Similarity 4.0%; Score 11; DB 5; Length 2527;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 59 AFKSAAEFT 69
 DB 1315 AFKSAAEFT 1325

RESULT 13

09XZB8 PRELIMINARY; PRT; 2710 AA.
 AC 09XZB8;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Variant-specific surface protein.
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN-CS2;
 MEDLINE-99238507; PubMed-10220443;
 RA Reeder J.C., Cowman A.F., Davern K.M., Beeson J.G., Thompson J.K.,
 RA Rogerson S.J., Brown G.V.;
 "The adhesion of Plasmodium falciparum-infected erythrocytes to
 RT chondroitin sulfate A is mediated by P. falciparum erythrocyte
 RT membrane protein 1.";
 RU Proc. Natl. Acad. Sci. U.S.A. 96:5198-5202(1999).
 DR EMBL; AF134154; AAD29126.1; -.
 DR InterPro; IPR004258; PFEMP.
 DR Pfam; PF03011; PFEMP. 2.
 FT NON_TER 2710
 FT 2710
 SQ SEQUENCE 2710 AA; 309247 MW; 0AC64F0A3D5BF512 CRC64;

Query Match
 Best Local Similarity 4.0%; Score 11; DB 5; Length 2710;
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRKLCV 38
 DB 941 CMPPRKLCV 951

RESULT 14

026032
ID Q26032 PRELIMINARY; PRT: 3006 AA.
AC Q26032;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Variant-specific surface protein.
CN VAR-3.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=95330813; PubMed=7606788;
RA Su X.Z., Heatwole V.M., Wertheimer S.P., Guinet F., Herfelde J.A.,
RA Peterson D.S., Ravetch J.A., Welles T.E.;
RT "The large diverse gene family var encodes proteins involved in
cytoadherence and antigenic variation of Plasmodium falciparum-
infected erythrocytes."
Cell 82:89-100(1995).
DR EMBL; L40609; AAA75397.1; -
DR InterPro: IPR004258; PFEMP.
DR Pfam: PF03011; PFEMP. 2
SO SEQUENCE 3006 AA; 343769 MW; 8FD5FF475F943C7A CRC64;

Query Match 4.0%; Score 11; DB 5; Length 3006;
Best Local Similarity 100.0%; Pred. No. 0.034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

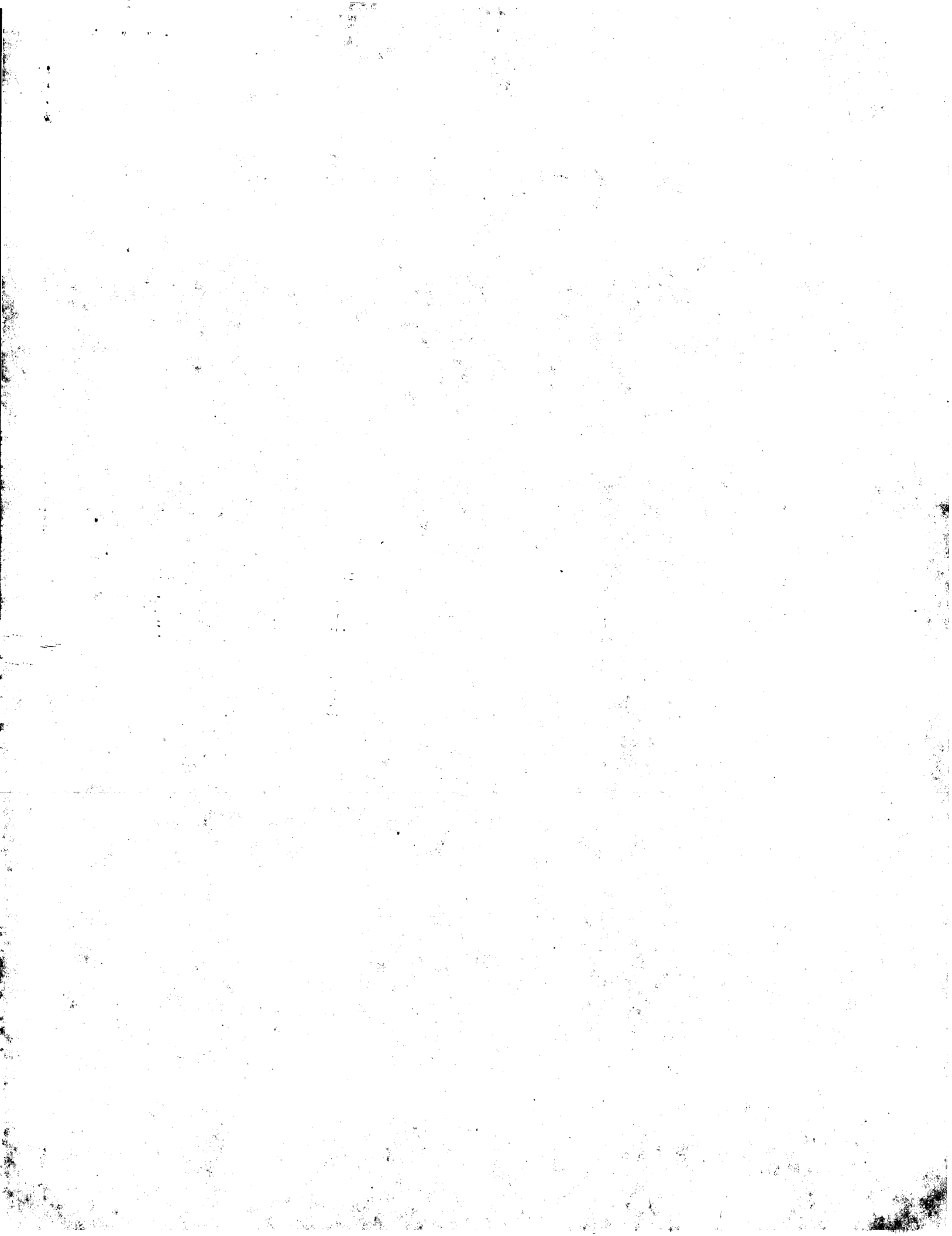
QY 28 CMPRRKRLCV 38
|||||
DB 1270 CMPRRKRLCV 1280

RESULT 15
Q9NFM8 PRELIMINARY; PRT: 81 AA.
ID Q9NFM8
AC Q9NFM8;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Var protein (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=21316723; PubMed=11424024;
RA Arley F., Hommel D., Le Scanf C., Duchemin J.B., Peneau C., Hulin A.,
RA Sarrhou J.L., Reynes J.M., Fandeur T., Mercereau-Pujalon O.;
RT "Association of severe malaria with a specific Plasmodium falciparum
RT genotype in French Guiana."
J. Infect. Dis. 184:237-241(2001).
DR EMBL; AJ27137; CAB86464.1; -
KW Malaria.
FT NON_TER 1
FT VARIANT 11 15 TFGDY -> IWRH.
FT NON_TER 81
FT NON_TER 81
SO SEQUENCE 81 AA; 9336 MW; 5849685D4A336C7A CRC64;

Query Match 3.6%; Score 10; DB 5; Length 81;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 102 MFTFGDYRD 111
|||||
DB 8 MFTFGDYRD 17

Search completed: April 28, 2003, 10:51:30
Job time : 29.333 secs



GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:49:21 ; Search time 16.8889 Seconds
(without alignments)
480.832 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPKKNSNGYPMQCGNIN.....KKEMIPQDKYKDERDKRRP 276

Scoring table: OLIGO
Gapop 60.0 , Gapect 60.0

Searched: 262574 seqs, 29422922 residues

rd size : 9

Total number of hits satisfying chosen parameters: 13

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	4.0	324	2 US-08-568-459A-17	Sequence 17, Appl
2	11	4.0	324	2 US-08-487-826B-29	Sequence 29, Appl
3	11	4.0	324	4 US-09-210-288-17	Sequence 17, Appl
4	11	4.0	921	2 US-08-568-459A-8	Sequence 8, Appl
5	11	4.0	921	2 US-08-487-826B-8	Sequence 8, Appl
6	11	4.0	921	2 US-08-568-459A-21	Sequence 21, Appl
7	10	3.6	311	4 US-08-487-826B-33	Sequence 33, Appl
8	10	3.6	311	4 US-09-210-288-21	Sequence 21, Appl
9	10	3.6	2710	2 US-08-568-459A-12	Sequence 12, Appl
10	10	3.6	2710	2 US-08-487-826B-12	Sequence 12, Appl
11	10	3.6	2710	4 US-09-210-288-12	Sequence 12, Appl
12	10	3.6	3060	2 US-08-487-826B-14	Sequence 14, Appl
13	10	3.6	3060	2 US-08-487-826B-14	Sequence 14, Appl

ALIGNMENTS

RESULT 1
US-08-568-459A-17
; Sequence 17, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/568,459A

FILING DATE: 07-DEC-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Israelien, Ned

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH121.001CP1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-0176

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 324 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: Internal

ORIGINAL SOURCE:

US-08-568-459A-17

Query Match 4.0%; Score 11; DB 2; Length 324;
Best Local Similarity 100.0%; Pred. No. 0.0034;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLTW 212
Db 204 RPOFLRWLTW 214

RESULT 2

US-08-487-826B-29
; Sequence 29, Application US/08487826B
; Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.

APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.

APPLICANT: Su, Xin-zhaun

APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knodde Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-08-487-826B-29

Query Match
Best Local Similarity 4.0%; Score 11; DB 2; Length 324;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 202 RPOFLRWLTEM 212
|||||
DB 204 RPOFLRWLTEM 214

RESULT 3
US-09-210-288-17
Sequence 17, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 324 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
US-09-210-288-17

Query Match
Best Local Similarity 4.0%; Score 11; DB 4; Length 324;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 202 RPOFLRWLTEM 212
|||||
DB 204 RPOFLRWLTEM 214

RESULT 4
US-08-568-459A-8
Sequence 8, Application US/08568459A
Patent No. 5849306

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-8

Query Match
Best Local Similarity 4.0%; Score 11; DB 2; Length 921;
Matches 100.0%; Pred. No. 0.0083;

TELEPHONE: (619) 235-85
TELEFAX: (619) 235-0176

ORIGINAL SC
TS-09-210-388-21

3.68; Score 10; DB 4; Length 311;

Best Local Similarity 100.0%; Pred. No. 0.033;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRKLC 37
Db 24 CMPPRKLC 33

RESULT 10
US-08-568-459A-12
Sequence 12, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum

US-08-568-459A-12
Query Match 3.6%; Score 10; DB 2; Length 2710;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRKLC 37
Db 1391 CMPPRKLC 1400

RESULT 11
US-08-487-826B-12
Sequence 12, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan

APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum

US-08-487-826B-12
Query Match 3.6%; Score 10; DB 2; Length 2710;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 28 CMPPRKLC 37
Db 1391 CMPPRKLC 1400

RESULT 12
US-09-210-288-12
Sequence 12, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-0176
; TELEFAX: (619) 235-8550
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2710 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; US-09-210-288-12

```

```

Query Match          3.6%; Score 10; DB 4; Length 2710;
Best Local Similarity 100.0%; Pred. No. 0.21;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 28 CMPPRQKLC 37
   |||||
Db 1391 CMPPRQKLC 1400

```

```

RESULT 13
US-08-487-826B-14
; Sequence 14, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Welliams, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelien, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH121.001CP1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176

```

```

; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3060 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-487-826B-14

```

```

Query Match          3.6%; Score 10; DB 2; Length 3060;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 28 CMPPRQKLC 37
   |||||
Db 1389 CMPPRQKLC 1398

```

```

Search completed: April 28, 2003, 10:53:15
Job time : 19.8889 secs

```


GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Computing Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:51:36 ; Search time 13.3333 Seconds
(without alignments)
1658.687 Million cell updates/sec

Title: US-10-087-013-2_COPY_1279_1554

Perfect score: 276
Sequence: 1 DCHPRKNSNGYPMQCNIN.....KKEMIDQKRYKDERDKRRF 276

Scoring table: OLIGO
Gapop 60.0 , Gapept 60.0

Searched: 301932 seqs, 80129803 residues

rd size : 9

Total number of hits satisfying chosen parameters: 4

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Published Applications AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	4.0	324	9	US-10-153-273-17
2	11	4.0	921	9	US-10-153-273-8
3	10	3.6	311	9	US-10-153-273-21
4	10	3.6	2710	9	US-10-153-273-12

ALIGNMENTS

RESULT 1
US-10-153-273-17

Sequence 17, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun

Willems, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb Martens Olson & Bear
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/153,273

FILING DATE: 21-May-2002

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/210,288

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Fuller, Michael

REGISTRATION NUMBER: 36,516

REFERENCE/DOCKET NUMBER: NIH121.1FWDV1

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:

LENGTH: 324 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHEICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: Internal

ORIGINAL SOURCE:

SEQUENCE DESCRIPTION: SEQ ID NO: 17:

US-10-153-273-17

Query Match 4.0%; Score 11; DB 9; Length 324;
Best Local Similarity 100.0%; Pred. No. 0.015;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 202 RPOFLRWLTEM 212

Db 204 RPOFLRWLTEM 214

RESULT 2
US-10-153-273-8

Sequence 8, Application US/10153273
Patent No. US20020169305A1

GENERAL INFORMATION:

APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Willems, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobb Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor

CITY: Newport Beach

STATE: California

COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-273-8

Query Match 4.0%; Score 11; DB 9; Length 921;
Best Local Similarity 100.0%; Pred. No. 0.035;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 202 RPOFLRWLTEM 212
Db 618 RPOFLRWLTEM 628

RESULT 3
US-10-153-273-21
Sequence 21, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:

APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 311 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
SEQUENCE DESCRIPTION: SEQ ID NO: 21:
US-10-153-273-21

Query Match 3.6%; Score 10; DB 9; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CMPPRQKLC 37
Db 24 CMPPRQKLC 33

RESULT 4
US-10-153-273-12
Sequence 12, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 US-10-153-273-12

Query Match

Best Local Similarity 3.68; Score 10; DB 9; Length 2710;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 28 CMPPRRQKLC 37
 |||||
 Db 1391 CMPPRRQKLC 1400

Search completed: April 28, 2003, 11:02:44
 Job time : 14.333 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 10.5249 Seconds
(without alignments)
4367.892 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 1918
Sequence: 1 TKEKQSYLSNDNKFVNIN.....CKDNTNACEFSHNTNP 345

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Minimum number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1918	100.0	3542	22 AAB62142	P. falciparum FCR3
2	446	22.3	2228	23 AAW93944	Plasmodium pfeum
3	432.5	22.5	2197	21 AAB18352	Plasmodium falcipa
4	431	22.5	2445	22 AAB6344	Malaria parasite
5	394.5	20.6	2703	16 AAR70236	P. falciparum Proj
6	394.5	20.6	2710	18 AAW22482	Plasmodium Proj3.
7	394.5	20.6	2710	21 AAY77904	P. falciparum Proj
8	394.5	20.6	3060	18 AAW22475	Plasmodium var-7.
9	394.5	20.6	3060	21 AAY77905	Plasmodium var-7.
10	380	19.8	700	16 AAR70235	P. falciparum EBL-

11	380	19.8	700	18 AAW22481	Plasmodium ebl-2.
12	380	19.8	700	21 AAY77903	P. falciparum ebl-
13	380	19.8	2182	18 AAW22476	Plasmodium var-1.
14	380	19.8	2182	21 AAY77906	Plasmodium var-1 P
15	357.5	18.6	1726	17 AAW00385	Truncated Plasmodi
16	357.5	18.6	2913	17 AAW00384	Plasmodium falcipa
17	231	12.0	793	16 AAR70234	P. falciparum E3ia
18	231	12.0	921	18 AAR70230	Plasmodium E3ia.
19	231	12.0	921	21 AAY77902	P. falciparum ebl-
20	162	8.4	807	21 AAB18311	Plasmodium falcipa
21	161	8.4	972	23 AAY76761	Plasmodium falcipa
22	153.5	8.0	696	21 AAB18181	Plasmodium falcipa
23	151.5	7.9	1086	23 AAY76760	Plasmodium falcipa
24	150.5	7.8	2380	21 AAB18315	Plasmodium falcipa
25	147.5	7.7	1516	21 AAB18195	Plasmodium falcipa
26	147	7.7	1421	23 AAY76764	Plasmodium falcipa
27	145	7.6	749	16 AAR70233	P. falciparum EBL-
28	145	7.6	749	18 AAW22479	Plasmodium ebl-1.
29	145	7.6	749	21 AAY77901	P. falciparum ebl-
30	144.5	7.5	1979	21 AAB18171	Plasmodium falcipa
31	143.5	7.5	1166	21 AAB18268	Plasmodium falcipa
32	142.5	7.4	1247	21 AAB18215	Plasmodium falcipa
33	142	7.4	3973	21 AAB18253	Plasmodium falcipa
34	140.5	7.3	1817	21 AAB18355	Plasmodium falcipa
35	137	7.1	1435	16 AAR70232	P. falciparum SABP
36	137	7.1	1435	18 AAW22477	Silic acid bindin
37	137	7.1	1435	21 AAY77900	P. falciparum SABP
38	137	7.1	1604	16 AAR70105	TNF-R-EBA 175 fusl
39	137	7.1	1786	14 AAR41043	CD4-EBA175 fusion
40	137	7.1	2010	21 AAB18218	Plasmodium falcipa
41	136	7.1	508	21 AAB18187	Plasmodium falcipa
42	136	7.1	1147	14 AAR41199	CAI antigen. Heli
43	136	7.1	1338	14 AAW80600	Helicobacter pylor
44	135.5	7.1	350	23 ABB07656	P. falciparum EBA-
45	135.5	7.1	1817	21 AAB18301	Plasmodium falcipa

ALIGNMENTS

RESULT 1
ID AAB62142 standard; Protein: 3542 AA.
XX AAB62142:
AC
XX
XX 29-MAY-2001 (first entry)
DT
XX
DE P. falciparum FCR3, varCSA protein.
XX
XX FCR3, varCSA protein; chondroitin sulfate A; CSA; var gene; PFPMP1;
KW erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
KW malaria; protozoacide.
XX
XX Plasmodium falciparum.
OS
XX
XX WO200116326-A2.
PN
XX
XX 08-MAR-2001.
PD
XX
XX 01-SEP-2000; 2000WO-US24195.
PE
XX
XX 01-SEP-1999; 99US-0152023.
PR
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
XX Gysin J, Pouvelle B, Fujii N, Smith J;
PI
XX WPI; 2001-235109/24.
XX N-PSDB; AAF57301.
DR
XX Novel FCR3, varCSA protein, useful for modulating parasitized red blood

PT cell binding, sequestration and onset of maternal malaria -
 XX
 PS Claim 12, Page 63-71; 78pp; English.
 XX
 CC The invention relates to a P. falciparum FCRI3 varCSA protein, that is
 CC capable of binding to chondroitin sulfate A (CSA). The var gene and the
 CC corresponding P. falciparum erythrocyte membrane protein 1 (PfEMP1)
 CC modulate adhesion of parasitized red blood cell (PRBC) to CSA. The
 CC protein and the encoding gene are useful for treating and preventing
 CC maternal malaria in a patient identified at a risk for contracting
 CC maternal malaria or in a patient afflicted with maternal malaria. The
 CC present sequence represents the P. falciparum FCRI3 varCSA protein.
 XX
 SQ Sequence 3542 AA:
 Query Match 100.0%; Score 1918; DB 22; Length 3542;
 Best Local Similarity 100.0%; Pred. No. 3e-138;
 Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 YEKELSYLSNDKPFVNNINSEYKOFYKELKETOYATNDFTLNLNEGKYGKGLPGEK 60
 403 YEKELSYLSNDKPFVNNINSEYKOFYKELKETOYATNDFTLNLNEGKYGKGLPGEK 462
 OY 61 DIFETMSADDKGIFRSEYCOVCPDCGVKCDGKITYHKSNDREYNNEDYKPPWGVKPT 120
 463 DIFETMSADDKGIFRSEYCOVCPDCGVKCDGKITYHKSNDREYNNEDYKPPWGVKPT 522
 OY 121 NITVLXSGNEQDITOKLEFNCSSSTNYKDKNNQKWEYKDEINRCKLEQTEINNDN 180
 523 NITVLXSGNEQDITOKLEFNCSSSTNYKDKNNQKWEYKDEINRCKLEQTEINNDN 582
 OY 181 PKTISHNFELMWTYLLDITKMNDKLTKCINNTTTCIDCNRNCLCFDRAVAKOKEE 240
 583 PKTISHNFELMWTYLLDITKMNDKLTKCINNTTTCIDCNRNCLCFDRAVAKOKEE 642
 OY 241 WNSIKKLFTRKKNNIOOSYSNNINLEGEYFVYMDKLDDEAKWKEIMENIRKKNEFSN 300
 643 WNSIKKLFTRKKNNIOOSYSNNINLEGEYFVYMDKLDDEAKWKEIMENIRKKNEFSN 702
 OY 301 LENNRYLENATEILLDHLKETATTCCKNNTNACETSHNATNP 345
 703 LENNRYLENATEILLDHLKETATTCCKNNTNACETSHNATNP 747
 DB
 RESULT 2
 AAM93944
 ID AAM93944 standard; protein: 2228 AA.
 XX
 AAM93944;
 30-JUN-1999 (first entry)
 DE P. falciparum PfEMP1 protein.
 XX
 KW Erythrocyte membrane protein; EMP; PfEMP1; malaria; antioclusional;
 KW glycosamino-glycan-like moiety; antiaggregational; antimalarial;
 KW antigen receptor; infected erythrocyte; rosette formation; blood cell;
 KW capillary occlusion; cerebral malaria; treatment; vaccine; detection;
 KW medicament; parasite; diagnosis; drug screening.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9915557-A1.
 PD 01-APR-1999.
 XX
 PE 18-SEP-1998; 98WO-SE01675.
 XX
 PR 19-SEP-1997; 97SE-0003386.
 XX
 PA (KARO-) KAROLINSKA INNOVATIONS AB.
 XX
 PI Barragan A, Carlson J, Fernandez V, Qijun C, Wahlgren M;

XX
 DR MPI: 1999-254692/21.
 XX
 PT New Isolated malaria polypeptides
 XX
 PS Claim 4, Page 67-74; 80pp; English.
 XX
 CC This invention describes a novel Plasmodium falciparum erythrocyte
 CC membrane protein (EMP), PfEMP1, which is capable of binding to a
 CC carbohydrate which exhibits at least one negatively charged
 CC glycosamino-glycan (GAG)-like moiety and has antiaggregational,
 CC antioclusional and antimalarial activity. The carbohydrates of the
 CC invention are capable of acting as receptors for malaria antigens
 CC present on the surfaces of malaria infected erythrocytes, by binding
 CC to these antigens the carbohydrates prevent rosette formation by the
 CC blood cells, this prevents occlusion of capillaries as is seen in
 CC cerebral malaria caused by Plasmodium falciparum. The products of the
 CC invention can be used to treat malaria or to vaccinate against it, or
 CC used to design a model to identify compounds that bind to PfEMP1. The
 CC carbohydrates, polypeptides and antibodies of the invention can be used
 CC as a medicament for dissolving the rosettes formed by erythrocytes
 CC infected by a malaria parasite. The products can also be used for
 CC detection, diagnosis and drug screening.
 XX
 SQ Sequence 2228 AA:
 Query Match 23.3%; Score 446; DB 20; Length 2228;
 Best Local Similarity 31.4%; Pred. No. 3.7e-25;
 Matches 131; Conservative 53; Mismatches 133; Indels 100; Gaps 20;
 OY 1 YEKELSYLSNDKPFVNNIN--SEYKOFYKELKETOYATNDFTLNLNEGKYGK---- 53
 415 YKEISDGGGRKRAVGGTTKYEYKSEFKKNDGIGYDAFLGLNNKACKDITDG 474
 OY 54 -----GGLPEKDTITNSADDKGIFRSEYCOVCPDCGVKCDGKITYH 97
 475 GKINKEVNSGGVGGSGGTSGASGIND--ENKGTFRSEYCOVCPDCGVK----H 526
 OY 98 KSDNREYNNEDYKPPG--VKPTN--ITVLXSGNEQD---ITOKLEFNC--NSS- 145
 527 KGNQWER-KTKYKMKRMSKLYKPRNGAMVLLSLAKYVAMTLKRWKEFLTQNSSD 585
 OY 146 -----TNYKDKNNQK-----WECYKDEINRCKLEQTEINND----- 179
 586 GSVGSVYVTGASGNSSEKELYDEMKC-YKNEVOYKNAVGEVEDDELKAGAGLITLP 644
 OY 180 NPK-----IISFHNPELMWTYLLDITKMNDK-LKTCINN-TTTHCI 220
 645 NPKNKEVSEAKSONNHADIOKTFHDEFFYVVAHMLKDSIMWTRKLSGISDGTWKR 704
 OY 221 DECNRNCLCFDRAVAKOKEEENSIKLEFTRKKNNIOOSY-SNINLEGEYEFK-----V 273
 705 NGCNKKCDCEFEKWKOKETEMKPKIDHFKTOEGIPGTYITTLILIKLOFLKEDTEENT 764
 OY 274 MDKLDDEAKWKEIMENIRKKNEFSNLE--NNRDYLENATEILLDHLKETATTCCKD 328
 765 ENSIDAEAEELKHLQKILKLENE--NNLAVNAGTEQKTLMDKLNLNELDARCKMD 820
 DB
 RESULT 3
 AAB18352
 ID AAB18352 standard; protein: 2197 AA.
 XX
 AC AAB18352;
 XX
 DT 07-NOV-2000 (first entry)
 XX
 DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:210.
 XX
 KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.
 XX
 OS Plasmodium falciparum.

XX XX WO200025728-A2.
 XX PN 11-MAY-2000.
 XX PD
 XX XX
 XX PF 05-NOV-1999; 99MO-US26796.
 XX PR 05-NOV-1998; 98US-0107131.
 XX PA (HOFF/) HOFFMAN S.
 XX PA (CARU/) CARUCI D.
 XX PA (GARD/) GARDNER M.
 XX PA (VENT/) VENTER J C.
 XX PI Hoffman S, Carucci D, Gardner M, Venter JC;
 XX DR WPI: 2000-365347/31.
 XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
 XX XX Plasmodium falciparum, useful as antimalarial vaccines and in the
 XX XX diagnosis of P. falciparum infection -
 XX XX
 XX XX Disclosure: Page 441-447; 577pp: English.
 XX PS
 CC CC The present invention describes proteins and their fragments (I) encoded
 CC CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
 CC CC vaccines against P. falciparum infection comprising (I) or (II).
 CC CC (I) and (II) are useful for the development of vaccines against
 CC CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC CC antibody raised to immunogens comprising the sequences of (I), are
 CC CC useful in the detection of infection with P. falciparum. Furthermore,
 CC CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC CC can aid the identification of drugs to treat or prevent P. falciparum
 CC CC infection, or they can be used to identify drug resistance in
 CC CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC CC subsequent identification of proteins encoded by it will help to expand
 CC CC our understanding of parasite biology, a process hampered by the
 CC CC complexity of the parasite lifecycle, and provide new targets for
 CC CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC CC resistance to insecticides have led to a resurgence of malaria in many
 CC CC parts of the world, and there is a pressing need for vaccines and new
 CC CC drugs. AA170078 to AA170287 and AA181844 to AA18352 represent nucleotide
 CC CC and protein sequences given in the present invention, but which are not
 CC CC specifically mentioned within the specification.
 CC CC
 CC XX
 CC XX Sequence 2197 AA;
 SQ
 Query Match 22.5%; Score 432.5; DB 21; Length 2197;
 Best Local Similarity 29.4%; Pred. No. 4e-24;
 Matches 116; Conservative 62; Mismatches 126; Indels 91; Gaps 14;
 OY 23 YKQFYEKLEKETOYATNDTFLNLNEGKCKGGLPGKEDTFTNSAD---DKGIFRSE 78
 Db 443 YKFFYELEKSEKGRKDDFLKLNEDVCKKIDKEKIDFTPAKNSNNEGTFTHSE 502
 OY 79 YQGVPCDGVKCDGKIKYTHSNDREVRNEDYKPPWGVKPTNTIVLYSGNEGSDITQKL 138
 Db 503 YKRCPCDGVKRYKRDQKDKYDCKTR--GKLYEPASGAGQTPKILKSGKQKEIFETKL 560
 OY 139 ENRCN---SSTN-----YKDKNNOK---WECYK-----DE 163
 Db 561 KAFEDQNGDTNSVANAGGADSGSKNSKELYEKKCKNEVOKVADKKNGEEDDEDE 620
 OY 164 NINR-----CKLE-----ONTEINNDNPKITSFNFELVATYLLRDIKKNDKLT 210
 Db 621 DVNDVKAAGGCTLENKKHESRNNSSNEPFOKTFHDFEFTWIGRPLNDSMTYKRGKYN 680
 OY 211 CINF--TTTHCIDECNRNCLCFDRWVKKQKEEWSNKKLFTKKNNIQOOSYSSNNLFE 269
 Db 681 CINPKRKCKNECKDCCGCKEWIKGKKKEWENIKKFKQEAFAKNNKRENSGIDMSG- 739
 OY 270 FFKVMDLID-----KDEAKMKELMENIKKKKNEFSNL----- 301

Db 740 ---LMDSDAVVLELLELEOLFQDIKDGVDVKELKIGKELDEKRRKKAEEAVVVVA 796
 OY 302 ENNRDYLENAIEILLDLKETATICKDNNTNEACE 336
 Db 797 DNQR---KTTIDKILQHEGDANNCLKTH--KEKCE 827
 RESULT 4
 ID AAB66344 standard; protein; 445 AA.
 XX AAB66344;
 AC AAB66344;
 XX
 DT 05-APR-2001 (first entry)
 XX
 DE Malarial parasite derived protein pf-EMP-1 fragment CIDR/A4.
 XX
 KW Malarial parasite; pf-EMP-1; vaccine; malaria; autoimmune disease;
 KW dendritic cell maturation; immune tolerance.
 OS
 OS Unidentified.
 XX
 XX WO200102005-A2.
 XX PN 11-JAN-2001.
 XX PD
 XX PF 30-JUN-2000; 2000MO-GB02546.
 XX PR 30-JUN-1999; 99GB-0015311.
 XX PA (ISIS-) ISIS INNOVATION LTD.
 XX PI Roberts DJ, Urban BC;
 XX XX WPI: 2001-138065/14.
 XX DR
 XX PT Treating mammalian dendritic cells in vitro to induce immune tolerance
 XX PT comprising exposing the cells to an agonist of the cell surface
 XX PT receptors CD36 and/or CD51, useful for treatment of autoimmune
 XX PT disorders -
 XX XX
 XX PS Claim 47; Fig 2; 148pp: English.
 XX XX
 CC CC The present invention describes methods of inducing immune tolerance in
 CC CC dendritic cells by exposing the cells to an agonist of the cell surface
 CC CC receptors CD36 or CD51, and of identifying molecules capable of
 CC CC preventing the adhesion of red blood cells infected with a malarial
 CC CC parasite to dendritic cells. The latter method involves the exposure of
 CC CC CD36 to the test molecule and parasitised human red blood cells, and
 CC CC determining the level of adherence of the blood cells to the dendritic
 CC CC cells. A molecule which causes a reduction in the adherence is useful in
 CC CC the treatment of malaria. The first method is useful in the treatment of
 CC CC autoimmune diseases.
 CC CC
 CC XX
 CC XX Sequence 445 AA;
 SQ
 Query Match 22.5%; Score 431; DB 22; Length 445;
 Best Local Similarity 28.1%; Pred. No. 6.3e-25;
 Matches 109; Conservative 71; Mismatches 134; Indels 74; Gaps 13;
 OY 11 NDNKFV---NNINSEYKQFYEKLEKETOYATNDTFLNLNEGKCKG---GLPGKEDI 62
 Db 7 NDDTTITTEYGTINNMRYKDEYKHLDE--KYTVDAFLNLNKRCKKNHPEVGGKKYI 65
 OY 63 TPTNSADKGFYSEYCOVPCDGVKCDGKIKYTHKSDNREVRN---EDYKPPW 115
 Db 66 DFNQNI---TFSHTEYREPCWGCIE-----EQDKKWRKINDSACKKEELVTPKE 115
 OY 116 GVPKPTNTIVLYSGNEGSDITQKLENFCSSTNYKDKNN-----OKWECYKDE- 163
 Db 116 NAKYTKINVLTSGGGHEDIARLKEFCIKYONGGGGSDCCGNSDSLCEBFWOCTOPDOL 175

QY 164 -----NINRCKLQNTPEINNDPKIISFHNFFELWYTLRDTIKW 204
 Db 176 EKVGSGEVDKLGAGGLCIPEKMKGEKVK -----KQTFNPFNFWAHVAKDSIDW 229
 QY 205 NDKLKTGIN-NTTHGIDCEGRNCLCFDRWYKQKEEEMNSIKLFTKKK-IOOSYSNI 262
 Db 230 RTOITKLSEKLLKCEKSGKSCNECEKRWKEKEKEMIKVQDFNQTDFLEKHYLV 289
 QY 263 NNLEFGYEFKVMKLDKDEAKWKELEMINIKRKNFEFNLNRRDYLFNAITELLDHLEK 322
 Db 290 ETILENYEFENIKAYGDLKSLIOEMKMKIKENKQNKRTD-----EDALDVLDFHEKEE 345
 QY 323 ATIC-----KDNNTNEACETSHNATNP 345
 Db 346 AEDCLDIHEDDDDDCEVELEKIPNP 373

RESULT 5
 AAR70236
 ID AAR70236 standard; Protein; 2703 AA.

AAR70236;
 DT 22-SEP-1995 (first entry)
 XX
 DE P. falciparum Proj3.
 XX
 KW Erythrocyte binding ligand; Proj3; binding domain; malaria; therapy;
 KW vaccine.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9507353-A.
 PD 16-MAR-1995.
 PF 07-SEP-1994; 94WO-US10230.
 PR 10-SEP-1993; 93US-0119677.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Chluis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 XX
 DR WPI: 1995-123427/16.
 DR N-PSDB; AAO83529.
 XX

New erythrocyte binding domain polypeptide(s) - isolated from Plasmodium binding proteins, used in diagnosis, treatment and prevention of malaria

Disclousure: Page 61-65; 81pp; English.

CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AAO83526), E3la (AAO83527), EBL-e2 (AAO83528) and Proj3
 CC (AAO83529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 XX
 XX
 SQ Sequence 2703 AA;

Query Match 20.6%; Score 394.5; DB 16; Length 2703;
 Best Local Similarity 26.2%; Pred. No. 4.4e-21;
 Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

QY 1 YEKEIOSY-----LSNNKFPVNNINSE-YKQFYEKLEKYOYANPFLINLNEBK 50
 Db 389 YDEIKKREKNAAGSGSRKQKRNAGGTTTNYGDEKFFYDELNKSEYRTVDFLEKLSNEE 448

QY 51 YC-----KGLPEKDDITFTNSA-----DDKGIIFYSEYQVCPDGVKCDGIKTYH 97
 Db 449 ICTKVADEEGGTTIDFKNVNSDSTSGASGTVNESQGFYFSKYCQPCPYCGVK-----KVNN 504
 QY 98 -KSDNDRERVNN-----EDYKPPWGVKPPNTIVLVSGNBOGDITQKLEPCH-----143
 Db 505 GGSNSNMEERNNKCKSGKIYEPKPKPEKGTITITLISGKHDDIEKLEKPFDERKNGDTI 564
 QY 144 -----SSTNKKDKNN-----QKWCYKNDENINRCKLQNTPEINNDPK-----11-- 184
 Db 565 NSGSGTGSGGGNSGRQELYEEMKC-YKGEVDVKGHDEDEDEYENKNNAGCITLKN 623
 QY 185 -----SPHNEFLWYTLRDTIKAKDKLCTI-NNTTGHC-ID 221
 Db 624 QKKNKEGGNTSEKPEDELOKTFNPFYVVAHMKDSITHMKKLOKRCLOKONRIRKCGNN 683
 QY 222 EGNRNCLEFDRWYKQKEEEMNSIKLFTKKNIQ-----QSYSNINLFE 267
 Db 684 KCNNDCECFKRWITQKKDKMGKIYQHF-KTONIKRGSGSDNTAELIPDHDYVLOYNIQE 742
 QY 268 GYF-----FKYMDKLDKDEAKWKELEMINIKRKNFEFNLNRRD-----YLEN 310
 Db 743 EFLKGDSEDASEKSENSIDAEAA-----BELKHLREIIESEDNNQSEASVGGVTEQKN 796
 QY 311 ATEILDLHLEKETATIC-----KDNNTNEACETSHNATNP 345
 Db 797 IMDKILNFKEDADLCLEIHEDEBEKEKGDGNECIEGENRYPN 842

RESULT 6
 AAW22482
 ID AAW22482 standard; Protein; 2710 AA.

AC AAW22482;
 XX
 DT 07-OCT-1997 (first entry)
 XX
 DE Plasmodium Proj3.
 XX
 KW DBL gene family; SAMP: stalic acid binding protein; vaccine; therapy;
 KW DABP; mezoizole; malaria; var-1; var-2; var-3; var-7; immune response;
 KW Plasmodium.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9640766-A2.
 PD 19-DEC-1996.
 PF 07-JUN-1996; 96WO-US09508.
 PR 07-JUN-1995; 95US-0487826.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Chluis C, Miller LH, Peterson DS, Slim KL, Su X;
 PI Wellens TE;
 XX
 DR WPI: 1997-052231/05.
 DR N-PSDB; AAT72897.
 XX

New malaria vaccines - contains cysteine-rich DBL family protein binding domains homologous domains of the Duffy and stalic acid binding proteins

Disclousure: Page 50-56; 96pp; English.

CC This sequence represents Proj3 of Plasmodium. Proj3 belongs to the
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stalic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of

CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).

CC Sequence 2710 AA:

Query Match 20.6%; Score 394.5; DB 18; Length 2710;

Best Local Similarity 26.2%; Pred. No. 4.4e-21; Mismatches 135; Indels 133; Gaps 20;

Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

1 YEKELISY-----LSNDKRVNNINSE--YKQPEYKLEKTOYATNDPFLNLNEKG 50
 DB 396 YDEIRKYEAGSGSGSRQKRDAGTTTNDGYEKRYDELKSEYTVKFLKLSNEE 455
 QY 51 YC-----KGLPEKIDITFTNSA-----DDKGIYRSYCYQVCPDCGKCGIRYTH 97
 DB 456 ICTRVKDEEGTIDFKVNSDSTSGASGTNYESGTFYRSKCYCPGCGYK-----RVNN 511
 QY 98 -KSDNREVRNN-----EDYKPPMGVYKPTNITVLYSGNBOGDTOKLENFNCN----- 143
 DB 512 GGSSENEBEKNNCKSGKLYEPKDEGTTITLKGCHDIEELKLFCDKKNDDTI 571
 QY 144 -----SSTNYKDKNN-----QKWECCYKDEINRCKLEONTETINNDPK-----IT-- 184
 DB 572 NSGSGGTGGSGGSGHQLYEEMKC-YKGEDVYKVGHDDEDEYEVNKNAGSLCLKN 630
 QY 185 -----SFNFEFLWYTYLLRTIKWNDKLCIT--NNTTTHC-ID 221
 DB 631 QKKRKEGGTSEKPEDEIOTKTFNPFYVVAHMLKDSIMHKKLOCLONGNRIKCGNN 690
 QY 222 ECNRNCICPDRWYKQKEEENSIKKLFTKKKNIO-----QSYYSNINLFE 267
 DB 691 KGNNDCECFKRWITQKDEMGKIVQHF-KTONIKRGSGSDNTAELIPDHDYVLYQYNIQ 749
 QY 268 GYF-----FKVMDKLKDEAKWKLMEINIKRKNESNLENND-----YLEN 310
 DB 750 EFLKGSDEASSEKSENSLDAEEA-----ELKHLREIIESEDNNOEASVGGVTDQKN 803
 QY 311 AIELLDHLKETATTC-----KDNNTNACETSHNATTNP 345
 DB 804 IMDKLWYERKDEADLCLEIHEDEEEREKDGKNCIEGENFRYNP 849

RESULT 7

AAV77904 standard; Protein; 2710 AA.

AAV77904:

13-JUN-2000 (first entry)

P. falciparum Proj3 binding domain polypeptide.

DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;

DABP; Salic Acid Binding Protein; SABP; malaria; vaccine; immunisation;

Plasmodium falciparum.

US5993827-A.

30-NOV-1999.

XX 07-JUN-1995; 95US-0487826.
 XX 10-SEP-1993; 93US-0119677.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Sim KL, Chlitis C, Peterson DS, Su X, Wellens TE, Miller LH;
 PI WPI: 2000-194198/17.
 DR N-PSDB; AA298286.

XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria

PS Disclosure: Columns 79-92: 93pp; English.

CC The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Salic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and salic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.
 CC Immunization with the polypeptide provides effective protection against
 CC malaria. The present sequence represents the Proj3 binding domain
 CC polypeptide.

XX Sequence 2710 AA:

Query Match 20.6%; Score 394.5; DB 21; Length 2710;

Best Local Similarity 26.2%; Pred. No. 4.4e-21; Mismatches 135; Indels 133; Gaps 20;

1 YEKELISY-----LSNDKRVNNINSE--YKQPEYKLEKTOYATNDPFLNLNEKG 50
 DB 396 YDEIRKYEAGSGSGSRQKRDAGTTTNDGYEKRYDELKSEYTVKFLKLSNEE 455
 QY 51 YC-----KGLPEKIDITFTNSA-----DDKGIYRSYCYQVCPDCGKCGIRYTH 97
 DB 456 ICTRVKDEEGTIDFKVNSDSTSGASGTNYESGTFYRSKCYCPGCGYK-----RVNN 511
 QY 98 -KSDNREVRNN-----EDYKPPMGVYKPTNITVLYSGNBOGDTOKLENFNCN----- 143
 DB 512 GGSSENEBEKNNCKSGKLYEPKDEGTTITLKGCHDIEELKLFCDKKNDDTI 571
 QY 144 -----SSTNYKDKNN-----QKWECCYKDEINRCKLEONTETINNDPK-----IT-- 184
 DB 572 NSGSGGTGGSGGSGHQLYEEMKC-YKGEDVYKVGHDDEDEYEVNKNAGSLCLKN 630
 QY 185 -----SFNFEFLWYTYLLRTIKWNDKLCIT--NNTTTHC-ID 221
 DB 631 QKKRKEGGTSEKPEDEIOTKTFNPFYVVAHMLKDSIMHKKLOCLONGNRIKCGNN 690
 QY 222 ECNRNCICPDRWYKQKEEENSIKKLFTKKKNIO-----QSYYSNINLFE 267
 DB 691 KGNNDCECFKRWITQKDEMGKIVQHF-KTONIKRGSGSDNTAELIPDHDYVLYQYNIQ 749
 QY 268 GYF-----FKVMDKLKDEAKWKLMEINIKRKNESNLENND-----YLEN 310
 DB 750 EFLKGSDEASSEKSENSLDAEEA-----ELKHLREIIESEDNNOEASVGGVTDQKN 803
 QY 311 AIELLDHLKETATTC-----KDNNTNACETSHNATTNP 345
 DB 804 IMDKLWYERKDEADLCLEIHEDEEEREKDGKNCIEGENFRYNP 849

RESULT 8

AAW22475 standard; Protein; 3060 AA.

Db 510 GGSNSMEWEKNNKGKCSKGLYPEPKDKEBTITITILKSGGHDIIEKLNKPEDENKGTI 569

QY 144 -----SSTNKDKNN-----QKMECYIKDENINRCKLEONTENINDPK-----IT-- 184

Db 570 NSGSGTGGSGGNGSGROBELYEEMC-YKGEYVAVKHGEDEDEEYEWKNAAGGLCIKN 628

QY 185 -----SEHNEFELMWYLLRDITIKMDKIKTCL-NNTTHC-ID 221

Db 629 OKKNEEGNTSEKPEDELOKTFNFFYVAHMLKDSIHMKKTLQRLQNGNRKICGN 688

QY 222 EGNRCLCFDRVVKQKEEEMWSIKLFTKKNIQ-----QSYYSININLFE 267

Db 689 KKNNDCECFKRWITQKDEMGKIVQH-KTQINRGSGSDMTALIPDHVYLVQYNQE 747

QY 268 GFE-----FKYMDLKDQEAKKWELMERIKKKKNEFSLENRP-----YLEN 310

Db 748 ELKQSDSEDASEKSENSLDAEEA-----BELKHLREIIESENNOASVGGVTEQKN 801

QY 311 ATEILLDLHKETATVTC-----KDNNTNEACFSSHATNP 345

Db 802 IMDKLTANKEDEADCLTEHDEEEKKGKGNCEIEGENFRNRP 847

RESULT 9

AA777905

XX ID AA777905 standard; Protein; 3060 AA.

XX AC

XX AA777905;

XX DT 13-JUN-2000 (first entry)

XX DE Plasmodium var-7 polypeptide.

XX KW DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation; protozoacide; var-7.

XX OS Plasmodium sp.

XX PN US5993827-A.

XX PD 30-NOV-1999.

XX PE 07-JUN-1995; 95US-0487826.

XX PR 10-SEP-1993; 93US-0119677.

XX PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX PI SIm KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;

XX DR WPI; 2000-194198/17.

XX DR N-PSDB; AAZ98287.

XX PT Isolated protein binding domains from Plasmodium vivax and Plasmodium falciparum erythrocyte binding proteins useful for vaccinating against malaria -

XX PS Disclosure; Columns 109-124; 93pp; English.

XX CC The invention relates to ebl-1 polypeptides that are encoded by the DBL (Duffy-binding like) gene family. The ebl-1 proteins are substantially identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid Binding Protein (SABP), which are soluble proteins that appear in the culture supernatant after erythrocytes infected with malaria release merozoites. Immunochemical studies indicate that DABP and SABP are the respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be used to vaccinate against malaria, especially caused by P. falciparum. Immunization with the polypeptide provides effective protection against malaria..The present sequence represents the var-7 polypeptide.

XX CC

XX Sequence 3060 AA;

Query Match 20.6%; Score 394.5; DB 21; Length 3060;
 Best Local Similarity 26.2%; Pred. No. 5.2e-21;
 Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

QY 1 YEKEIOSY-----LSNDKFNANNINSE-YKQFYEKLTQYATNDFTLNECK 50
 DB 394 YDEIKKYENGASGSRORADAGITTTTYDGEKFFYDELNKSEYRTYDKLEKLSNEE 453
 QY 51 YC-----KGLPGKEDITFTNSA-----DDKGIYRSEYCOVCPDGVKCDGIRK 97
 DB 454 ICTKVKDESGGTIDFNKVNSDSTSGASGTNVEGQTFYRSKYQPCPGYGVK-----KVN 509
 QY 98 -KSDNDRERVNN-----EDYKPPMGVCKPNTITVLYSGNEQGITQKLEFCN----- 143
 DB 510 GGSNNMEERNKCKCKSGKLYEPKPDKEGTTITLKSQGHDIIEKLNKFCDEKNDFTI 569
 QY 144 -----SSTNFKDKNN-----QKWEQYKDEINNRCKLEQNTFINDNPR----- 154
 DB 570 NSGSGSTGGSGGSGRQELIYEMKC-YKGEDVYKVGHDDEDEDEYENVNNAGGLCTLKN 628
 DB 185 -----SFHNEFELWVLYLDRITKMDKLKTCI-NNTTTC-ID 221
 DB 629 QKKNKEEGNTSEKPEDELQKTFNPFYVVAHMLKDSIMKKKLRCLONGNRKICGN 688
 QY 222 ECRNRCICDRWYKQKEEEMNSIKLFTYKKKNIQ-----QSYININNLFE 267
 DB 689 KCNNDCECFRMTOKRKDEGKIYQHF-KTONIKRGSGSDNTAELLPEFDHYLYQYMLQE 747
 QY 268 GYF-----FKYMDKLDKDEAKKELMENIKRKKNEHSNLENNRD-----YLEN 310
 DB 748 EFKKSGSEDAESEKSENSIDAEEA-----EELKHLMEITIESEDNNDEASVGGVTEQKN 801
 QY 311 ALELLDLHLEKPTATIC-----KDNNTNEACETSHNATTNP 345
 DB 802 IMOKLTNYEKDEADLCLEIHEDEEKEKGDNECIEEGENFRYNP 847

RESULT 10
 AAR70235
 ID AAR70235 standard; Protein; 700 AA.
 AC AAR70235;
 XX
 DT 22-SEP-1995 (first entry)
 XX
 DE P. falciparum EBL-e2.
 Erythrocyte binding ligand; EBL-e2; binding domain; malaria; therapy;
 vaccine.
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9507353-A.
 PD 16-MAR-1995.
 XX
 PF 07-SEP-1994; 94WO-US10230.
 XX
 PR 10-SEP-1993; 93US-0119677.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 XX
 DR WPI: 1995-123427/16.
 DR N-PSDB; AAO83528.
 XX
 PT New erythrocyte binding domain polypeptide(s) - isolated from
 PT Plasmodium binding proteins, used in diagnosis, treatment and
 PT prevention of malaria
 XX

PS Disclosure; Page 56-57; 81pp; English.
 CC Erythrocyte binding ligand (EBL) family genes were cloned from
 CC P. falciparum chromosome 7 subsegment libraries constructed during
 CC genetic studies of the chloroquine resistance locus. The 4 genes,
 CC EBL-e1 (AA083526), E3ia (AA083527), EBL-e2 (AA083528) and Pro13
 CC (AA083529), encode the proteins given in AAR70233-36, respectively. The
 CC binding domains of such proteins can be expressed e.g. in E. coli,
 CC yeast, mammalian, insect, and in vaccinia virus and adenovirus-infected
 CC cells, and provide protection against P. falciparum.
 XX

SO Sequence 700 AA:

Query Match 19.8%; Score 380; DB 16; Length 700;
 Best Local Similarity 28.7%; Pred. No. 9.7e-21;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

QY 1 YEKEI-----QSYLSNDKFNANNINSEYKQFYEKLTQYATNDFTL 43
 DB 285 YETELISGGSGKSPKRTKRAARSSSSDDN-----GYESKFKKLEKVGQYDVDFKL 336
 QY 44 NLNNEGYCKG-----GLPGKEDITFTNSADOKGIYRSEYCOVCPDGVKCDGIRK 99
 DB 337 KIINKESICQKQOVGNKADNDFTNEKYK-TFSKTEICEPCPGCLEKGPWKYKG 395
 QY 100 DNDREYNNEDYKPPMGVCKPNTITVLYSGNEQGITQKLEFCNSSTNYKDKNNQKECY 159
 DB 396 DKTCGSAKRTYDPK---NITDIPVLYPDKSQQNLKRYKNEFCRGAP-GGGQIKKKQCY 451
 QY 160 YKD-----ENINCKLEQNTFINDNKKIISFHNEFLWVLYLDRITKMDKLKTCI 214
 DB 452 YDERPSSKNNNNKCEGTWTKFTQKQYKSYNVEFMDWHMDHDSVEKTELSKIN 511
 QY 215 TTT-----HCIDECNRNCLCFDRWYKQKEEEMNSIKLFTYKKKNIQSYVINNLFE 270
 DB 512 NTNGNTRNNNNKCKTDGCGQKWEKKQEMAKDHFGQTDYVQ-----KGLIYFSP 566
 QY 271 FRYMDKLDKDEAKKELMENIKRKKNEHSNLENNRDYLE-----NALELL 316
 DB 567 YGVLDIV-----LKGNNLQNIKDVHGDITDIDIKHKKLLDEDAVAVALGKDNFTIDKL 622
 QY 317 DILKETATICKDNNTNFACE 336
 DB 623 QHEKEQAQCKOR--QEECE 640

RESULT 11
 AAW22481
 ID AAW22481 standard; Protein; 700 AA.
 AC AAW22481;
 XX
 DT 07-OCT-1997 (first entry).
 XX
 DE Plasmodium ebl-2.
 XX
 KW DBL gene family; SABP; sialic acid binding protein; vaccine; therapy;
 KW Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KW DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 XX
 OS Plasmodium falciparum.
 XX
 PN WO9640766-A2.
 PD 19-DEC-1996.
 XX
 PF 07-JUN-1996; 96WO-US09508.
 XX
 PR 07-JUN-1995; 95US-0487826.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX

PI Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 DR WPI; 1997-052231/05.
 DR N-PSDB; AAT72896.
 XX
 PT New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and sialic acid
 PT binding proteins
 PS
 PS Disclosure: Page 46-48; 96pp; English.

This sequence represents ebl-2 of Plasmodium. Ebl-2 belongs to
 the Duffy binding like (DBL) family of genes which have homology to the
 Duffy antigen binding protein (DABP) and sialic acid binding protein
 (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 var family of genes modulate cytoadherence and antigenic variation of
 Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 protein (DABP) are soluble proteins that appear in the culture
 supernatant after infected erythrocytes release merozoites. DABP and
 SABP mediate the binding of merozoites and schizonts to the erythrocyte
 surface. These proteins are necessary for erythrocyte invasion by the
 parasite. This sequence can be used in the compositions of the
 invention. The compositions are for the treatment and prevention of
 malaria, and comprise either a nucleotide sequence or encoded polypeptide
 of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 family of genes having homology with conserved regions of DABP and SABP.
 The compositions are used for the treatment and prevention of malaria.
 They are also used in the preparation of vaccines for inducing a
 protective immune response in a mammal to Plasmodium merozoites
 (especially Plasmodium falciparum or Plasmodium vivax).

Sequence 700 AA:

Query Match 19.8%; Score 380; DB 18; Length 700;
 Best Local Similarity 28.7%; Pred. No. 9.7e-21;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

1 YKEI-----QYLSNDNKRNVNINSERYKQFEKLEKETOYATNDTFL 43
 285 YETELISGGSGSPKRTKRAARSSSSDDN-----GESKRYKKLEKGYQDVDFKL 336
 44 NLNLEKRYCKG---GLPGEKDTFTNSADDKGIFYRSEYQVCPDGCYKCGIKRYTHKS 99
 337 KILNKEGICQKQPOYGENEKADNVDFTNEKYK-TFSRTLEICPCWCGLEKGGPPWVKYK 395
 100 DNDREVRNNEDEKPPWGVKPTNITVLYSGNEQGITOKLEFNCSSSTNYKDKNNQWECY 159
 396 DKTGSAKRTYDPR---NITDIPVLYPDKSQNLIKRYKNCCEKGP-GGGQIRKWCY 451
 160 YKD-----ENINRCKLEQNTTEINNDNPKIISPHNEFELWVYLLRDTIKWMDKLKTCINN 214
 452 YDEHRPSSKNNNNNCYEGTWDKFTQKQYKSYNVEFMDVHDLHDSVEWKTLSKCIINN 511
 215 TTT-----HCIDECNRNCLCFDRWYKQKEEWNISIKLFTKKNIQOASYISNINLFEGYF 270
 512 NTNGNTCRNNKCKTDCGCFQKWEKKQOEWMALIKHFGKQTDIYQ-----KGLIVFSP 566
 271 FKVMQKLDKDEAKWELMENIKRKKNEFSNLENNRPLYE-----NAIELLL 316
 567 YGVLDLV---LKGNNLQNIKIDVHGDITDIKIKLLDEDAVAVVLGKDNNTTIDKLL 622
 317 DILKETATICKDNNTNEACE 336
 623 QHEKQAEQCKOK--QECE 640

RESULT 12

AA77903
 ID AA77903 standard; Protein; 700 AA.
 AC AA77903;
 XX

DT 13-JUN-2000 (first entry)

XX P. falciparum ebl-2 polypeptide.

XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KW DABP; Sialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 KM protozoacide.
 XX

XX Plasmodium falciparum.

XX US593827-A.

XX 30-NOV-1999.

XX 07-JUN-1995; 95US-0487826.

XX 10-SEP-1993; 93US-0119677.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;

XX WPI; 2000-194198/17.

XX N-PSDB; AAZ98285.

XX Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -

XX Disclosure: Columns 67-72; 93pp; English.

The invention relates to ebl-1 polypeptides that are encoded by the DBL
 (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 identical to the Duffy Antigen Binding Protein (DABP) and Sialic Acid
 Binding Protein (SABP), which are soluble proteins that appear in the
 culture supernatant after erythrocytes infected with malaria release
 merozoites. Immunochemical studies indicate that DABP and SABP are the
 respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 and sialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 used to vaccinate against malaria, especially caused by P. falciparum.
 Immunization with the polypeptide provides effective protection against
 malaria. The present sequence represents the ebl-2 polypeptide.

Sequence 700 AA:

Query Match 19.8%; Score 380; DB 21; Length 700;
 Best Local Similarity 28.7%; Pred. No. 9.7e-21;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

1 YKEI-----QYLSNDNKRNVNINSERYKQFEKLEKETOYATNDTFL 43
 285 YETELISGGSGSPKRTKRAARSSSSDDN-----GESKRYKKLEKGYQDVDFKL 336
 44 NLNLEKRYCKG---GLPGEKDTFTNSADDKGIFYRSEYQVCPDGCYKCGIKRYTHKS 99
 337 KILNKEGICQKQPOYGENEKADNVDFTNEKYK-TFSRTLEICPCWCGLEKGGPPWVKYK 395
 100 DNDREVRNNEDEKPPWGVKPTNITVLYSGNEQGITOKLEFNCSSSTNYKDKNNQWECY 159
 396 DKTGSAKRTYDPR---NITDIPVLYPDKSQNLIKRYKNCCEKGP-GGGQIRKWCY 451
 160 YKD-----ENINRCKLEQNTTEINNDNPKIISPHNEFELWVYLLRDTIKWMDKLKTCINN 214
 452 YDEHRPSSKNNNNNCYEGTWDKFTQKQYKSYNVEFMDVHDLHDSVEWKTLSKCIINN 511
 215 TTT-----HCIDECNRNCLCFDRWYKQKEEWNISIKLFTKKNIQOASYISNINLFEGYF 270
 512 NTNGNTCRNNKCKTDCGCFQKWEKKQOEWMALIKHFGKQTDIYQ-----KGLIVFSP 566
 271 FKVMQKLDKDEAKWELMENIKRKKNEFSNLENNRPLYE-----NAIELLL 316
 567 YGVLDLV---LKGNNLQNIKIDVHGDITDIKIKLLDEDAVAVVLGKDNNTTIDKLL 622

OY 317 DHLEKATATICKDNNTNEACE 336
 DB 623 QHEKEQAECKOK--OECE 640
 RESULT 13
 ID AAM22476 standard; Protein; 2182 AA.
 XX AAM22476;
 AC 12-SEP-1997 (first entry)
 DE Plasmodium var-1.
 XX DBL gene family; SABP; stialic acid binding protein; vaccine; therapy;
 KM Duffy binding like gene; Duffy antigen binding protein; erythrocyte;
 KM DABP; merozoite; malaria; var-1; var-2; var-3; var-7; immune response;
 KM Plasmodium.
 XX Plasmodium vivax.
 XX Plasmodium falciparum.
 XX WO9640766-A2.
 XX 19-DEC-1996.
 XX 07-JUN-1996; 96WO-US09508.
 XX 07-JUN-1995; 95US-0487826.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX Chitnis C, Miller LH, Peterson DS, Sim KL, Su X;
 PI Wellens TE;
 XX WPI: 1997-052231/05.
 DR N-PSDB; AAT72883.
 XX New malaria vaccines - contains cysteine-rich DBL family protein
 PT binding domains homologous domains of the Duffy and stialic acid
 PT binding proteins
 XX Claim 8: Page 70-74; 96pp; English.
 XX This sequence represents var-1 of Plasmodium. Var-1 belongs to
 CC the Duffy binding like (DBL) family of genes which have homology to the
 CC Duffy antigen binding protein (DABP) and stialic acid binding protein
 CC (SABP) conserved regions (see AAT72889 and AAT72888 respectively). The
 CC var family of genes modulate cytoadherence and antigenic variation of
 CC Plasmodium infected erythrocytes. SABP and the Duffy antigen binding
 CC protein (DABP) are soluble proteins that appear in the culture
 CC supernatant after infected erythrocytes release merozoites. DABP and
 CC SABP mediate the binding of merozoites and schizonts to the erythrocyte
 CC surface. These proteins are necessary for erythrocyte invasion by the
 CC parasite. This sequence can be used in the compositions of the
 CC invention. The compositions are for the treatment and prevention of
 CC malaria, and comprise either a nucleotide sequence or encoded polypeptide
 CC of the var-1, var-2, var-3 or var-7 genes of the DBL gene family, a
 CC family of genes having homology with conserved regions of DABP and SABP.
 CC The compositions are used for the treatment and prevention of malaria.
 CC They are also used in the preparation of vaccines for inducing a
 CC protective immune response in a mammal to Plasmodium merozoites
 CC (especially Plasmodium falciparum or Plasmodium vivax).
 CC
 SO Sequence 2182 AA;
 Query Match 19.8%; Score 380; DB 18; Length 2182;
 Best Local Similarity 28.7%; Pred. No. 4.3e-20;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;
 OY 1 YEKI-----OSLSNDNKEFVNINSEYKQYKLEKNOVATNTEFL 43
 ||||| | :||| ||| | |||

DB 413 YETISGGGSKSPKRRKRAARSSSSSDN-----GYSKTYKKLEKYGVDVDFKL 464
 OY 44 NLNEGAYCKG---GLPEKNDITFTNSADDKGIFRSEYCOVCPDGVCKDGIRYTHKS 99
 DB 465 KILNKGICQKQPOVGNKADNVDFTEKRYK-TFSRTEICEPCPWCGLKGPWPKYK 523
 OY 100 DNDREYVNNEDYKPPWGVKPINITYLVISGNEQGITQKLENFCSNSTYKDKNNQMECY 159
 DB 524 DKTCGSAKTKYDPR---NITDIPVLYPDKSQONILKRYKNFCERKAP-GGQIKKMCY 579
 OY 160 YKD-----ENINRCKLDEQNTFINNDNPKIISFNHFEFLMYTYLDRITKMDKLTCTNN 214
 DB 580 YDEHRPSSKNNNCVEGTWDFETQKQTKSYNVFMDVMDHMLDSVEMTELSKCTNN 639
 OY 215 TTT-----HCIDECNRNCLCFDRWVKOKEEENSIKLTFTKKNIOQSYSTNNLPEGYF 270
 DB 640 NTNNTCRNNKCKTDCGCPKWEKQEQEMALKDHGKQDTLIVQ-----KGLIVSP 694
 OY 271 FKVDKLDKDEAKKKELEMTENIKRRKNEFSNLENNRDYLE-----NAIEELL 316
 DB 695 YGVLDLV---LKGGNLQNIKDVHGDPTDIIKIKLLDEEDAVAVVLGKDNFTIDRL 750
 OY 317 DHLEKATATICKDNNTNEACE 336
 DB 751 QHEKEQAECKOK--OECE 768
 RESULT 14
 ID AAY77906 standard; Protein; 2182 AA.
 XX AAY77906;
 XX 13-JUN-2000 (first entry)
 DE Plasmodium var-1 polypeptide.
 XX DBL gene; Duffy-binding like gene; ebl-1; Duffy Antigen Binding Protein;
 KM DABP; Stialic Acid Binding Protein; SABP; malaria; vaccine; immunisation;
 KM protozoacide; var-1.
 XX Plasmodium sp.
 OS US5993827-A.
 XX 30-NOV-1999.
 XX 07-JUN-1995; 95US-0487826.
 XX 10-SEP-1993; 93US-0119677.
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PI Sim KL, Chitnis C, Peterson DS, Su X, Wellens TE, Miller LH;
 DR WPI: 2000-194198/17.
 DR N-PSDB; AAZ98288.
 PT Isolated protein binding domains from Plasmodium vivax and Plasmodium
 PT falciparum erythrocyte binding proteins useful for vaccinating against
 PT malaria -
 XX Disclosure; Columns 129-140; 93pp; English.
 XX The invention relates to ebl-1 polypeptides that are encoded by the DBL
 CC (Duffy-binding like) gene family. The ebl-1 proteins are substantially
 CC identical to the Duffy Antigen Binding Protein (DABP) and Stialic Acid
 CC Binding Protein (SABP), which are soluble proteins that appear in the
 CC culture supernatant after erythrocytes infected with malaria release
 CC merozoites. Immunohistochemical studies indicate that DABP and SABP are the
 CC respective ligands for Plasmodium vivax and Plasmodium falciparum Duffy
 CC and stialic acid receptors on erythrocytes. The ebl-1 polypeptides may be
 CC used to vaccinate against malaria, especially caused by P. falciparum.

CC Immunization with the polypeptide provides effective protection against
CC malaria. The present sequence represents the var-1 polypeptide.

SO Sequence 2182 AA;

Query Match 19.8%; Score 380; DB 21; Length 2182;
Best Local Similarity 28.7%; Pred. No. 4.3e-20;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

```
QY 1 YKEKEI-----QSLSDNKKFVNNINSEYKQFEKLEKQYATNDPEL 43
  |||||
DB 413 YETELSGGSGSKSPKRTKRAANSSSSSDN-----GYESKFTYKLEKVGTYDYDKFL 464
QY 44 NLNKGKCKG---GLPGEKDIFTNSADKGFYRSEYSCVPCDGVKCDGKITYTHS 99
  |||||
DB 465 KLNKGKCKGKQYQNGEKADNVDFTEKRYK-TFSRTEICPCPCWGLGKGGPPMKVKG 523
QY 100 DNDREVRNNEDEKPPWGVKFTNITVLYSGNEGDTQKLENCSSTNKKDNQKMEY 159
  |||||
DB 524 DKTCGSARTKTYDPK---NITDIPVLPKRSQONTLKKRNKCEKGP-GGGQIKRMOCY 579
  |||||
DB 160 YKD-----ENIRCKLEQNTLEINNDNPKIISPHNFELWVTLRDTIKMNDKLTQICIN 214
  |||||
DB 580 YDEHPSSKNNNNCVGEWTDKFTQKQYKSTNVEFWMDVHMLHDSYEMKTELKCIIN 639
QY 215 TTT-----HCIDECNRNCLCFEDRWKQKEEWSNKKLFTRKKNIOQSYYSNINLEEGYF 270
  |||||
DB 640 NTNGWTCRRNNCKTDCGCFQKVEKQKQEWMAIKDHFGKQNDIYQO-----KGLIVESP 694
QY 271 FRVMDKLDKDEKMKELMENIKRKNERSNLENNRDYLE-----NAIEILL 316
  |||||
DB 695 YGVLDLV-----LKGNLQNIKIDVGHDDIDIKIKLLDEDDAVAVVLGKNDNTTIDKLL 750
QY 317 DLKKEFATICKDNTNEACE 336
  |||||
DB 751 QHEKEQAECKQK--QECE 768
```

RESULT 15
ID AAM00385.
AAM00385 standard; Protein; 1726 AA.

AC AAM00385;
DT 21-FEB-1997 (first entry)
XX
DE Truncated Plasmodium falciparum erythrocyte membrane protein.
XX
Plasmodium falciparum; erythrocyte membrane protein; malaria;
detection; identification; treatment; prevention; parasite.

OS Plasmodium falciparum MC type.
XX
FH Key Location/Qualifiers
FT Domain 62..394
FT /label= Duffy binding ligand domain 1
FT Region 607..648
FT /note= "Cysteine rich motif"
FT Domain 839..1282
FT /label= Duffy binding ligand domain 2
FT 1488..1523
FT /note= "Cysteine rich motif"
XX
XX MO963736-A1.
XX
XX 31-OCT-1996.
XX
XX 26-APR-1996; 96WO-US05798.
XX
XX 27-APR-1995; 95US-0430908.
XX
XX (AFYF-) AFFYMAX TECHNOLOGIES NV.

PI Baruch DI, Howard RJ, Pasloske BL;
XX
XX WPI: 1996-497376/49.
DR N-PSDB; AAT41853.

PT New Plasmodium falciparum erythrocyte membrane proteins - used to
CC develop products for the diagnosis, treatment or prevention of
PT malaria parasite infections
XX
XX
XX
PS Claim 1; Figure 12; 149pp; English.

CC A polypeptide comprising a Plasmodium falciparum (Pf) erythrocyte
CC membrane protein 1 (PfEMP1) or active fragments or analogues of that
CC protein can be used in the treatment or prevention of symptoms of a
CC malaria parasite infection. The polypeptides can inhibit, block or
CC reverse the sequestration of erythrocytes in patients suffering from
CC malaria. Nucleic acids derived from the PfEMP1 gene can be used as
CC probes and primers to identify a Plasmodium falciparum parasite, the
CC primers used to generate characteristic amplification patterns from
CC different P. falciparum strains. Antibodies specifically
CC immunoreactive with the PfEMP1 polypeptide or its fragments may be
CC used in diagnosis of malaria infection. This is a truncated PfEMP1
CC protein of the MC type of Plasmodium falciparum. The full length
CC PfEMP1 protein is given in AAM00384.

SO Sequence 1726 AA;

Query Match 18.6%; Score 357.5; DB 17; Length 1726;
Best Local Similarity 27.4%; Pred. No. 1.7e-18;
Matches 104; Conservative 64; Mismatches 137; Indels 75; Gaps 16;

```
QY 1 YKEKEI-----QSLSDNKKFVNNINSEYKQFEKLEKQYATNDPELNLNKGKCK--G 54
  |||||
DB 394 YDEINLKTHTGTTITGNGKINNL--YGHFYKILKK-YTPVDSLOKLNDEALCKKRP 449
QY 55 GLPGEKDIFTNSADKGFYRSEYSCVPCDGVKCD-----GKITYTHS-DNDREVRNNE 109
  |||||
DB 450 NVGNKASTVDENNEVNTTFSHTTYCEACPCWGAQKKNKGKAKKESCAKKEKREIFNK 509
QY 110 DYKPPWGVKFTNITVLYSGNEGDTQKLENCSSTNKKDNQKMECYKDEINRCK 169
  |||||
DB 510 E-----NSTDIKILTPKGRSKTLEKTKTCXQKGRK---NDIWKCHYDNG----- 554
QY 170 LEQNTLEINND-----NPKIISPHNFELWVTLRDTIKMNDKLTQICIN-NTT 217
  |||||
DB 555 TDDQTFDSDNDVYLDGWNLTKEKIMSYNAFPMWVHMLIDSIRKWDHGRICINDKGK 614
QY 218 HCIDECNRNCLCFEDRWKQKEEWSNKKLFTRKKNIOQSYYSNINLEEGYFVMDKL 277
  |||||
DB 615 TCIKGNKRCICFQKVEKQKTEWGIKIDHFRKQKIDPQDTHD-----DFIQTL--L 665
QY 278 DKDEAKMKELMENIKRKNERSNLE-----NNRDYL-----ENAIELL 315
  |||||
DB 666 MKD--LLEETIDYGDANEIKRIEALLEQAGVGVGIDFAALAGLYTKGVAKDITIDKL 723
QY 316 DLKKEFATICKDNTNEAC 335
  |||||
DB 724 LQHEQKADKCLKTHDTC 743
```

Search completed: April 28, 2003, 10:30:04
Job time : 17.5249 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 6.38122 Seconds
(without alignments)
5197.501 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 1918
Sequence: 1 YKEIQSYLSNDNKFVNIN.....CKDNTNEACETSHNATNP 345

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	771	40.2	3006	2 T28625	variant-specific s
2	501.5	26.1	2647	2 T28161	hypothetical prote
3	499.5	26.0	2212	2 T28157	erythrocyte membra
4	457	23.8	2664	2 T28626	variant-specific s
5	446	23.3	2228	2 T14029	variant-specific s
6	437	22.8	3026	2 T28431	variant surface pr
7	432.5	22.5	2197	2 B71600	variant-specific s
8	403.5	21.0	2706	2 T28155	variant-specific s
9	394.5	20.6	3078	2 T28432	variant-specific s
10	380	19.8	2182	2 T28634	variant-specific s
11	363	18.9	1711	2 C71625	variant-specific s
12	361.5	18.8	2135	2 T14602	variant-specific s
13	357.5	18.6	1729	2 T18396	erythrocyte membra
14	357.5	18.6	1729	2 T18396	variant-specific s
15	331	17.3	2042	2 T18399	variant-specific s
16	162	8.4	807	2 B71605	hypothetical prote
17	155.5	8.1	2265	2 T28627	rhodopy protein
18	153.5	8.0	686	2 E71620	hypothetical prote
19	150.5	7.8	2380	2 E71604	hypothetical prote
20	150	7.8	4550	2 T18440	hypothetical prote
21	149	7.8	1005	2 A64465	hypothetical prote
22	148.5	7.7	3394	2 T18501	hypothetical prote
23	148.5	7.7	3724	2 T18427	hypothetical prote
24	147.5	7.7	1516	2 E71619	RAD2 endonuclease
25	146	7.6	1939	2 T18372	repeat organellar
26	146	7.6	2339	2 A45597	DNA-directed RNA p
27	145.5	7.5	407	2 C45600	asparagine-rich bl
28	144.5	7.5	1979	2 C71622	hypothetical prote
29	143.5	7.5	1166	2 H71609	hypothetical prote

30	142.5	7.4	1247	2 E71616	hypothetical prote
31	142	7.4	1711	2 T18429	hypothetical prote
32	140.5	7.3	1817	2 H71611	probable secreted
33	140	7.3	2136	2 A05037	hypothetical prote
34	139	7.2	2401	2 T28676	hypothetical prote
35	138	7.2	1127	2 T28317	rhodopy protein
36	137	7.1	1435	2 A37793	ORF MSV156 hypote
37	137	7.1	2010	2 B71616	erythrocyte-bindin
38	137	7.1	3973	2 B71612	phosphatase (acid
39	136	7.1	508	2 E71620	hypothetical prote
40	136	7.1	911	2 T18451	hypothetical prote
41	136	7.1	1215	2 B48281	hypothetical prote
42	136	7.1	4961	2 T18489	cytochrome associat
43	135.5	7.1	1817	2 D71606	hypothetical prote
44	135.5	7.1	2657	2 T18497	hypothetical prote
45	135	7.0	2510	2 T28160	hypothetical prote

ALIGNMENTS

RESULT 1
T28625
variant-specific surface protein 3 - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T28625
R/Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, Cell 82, 89-100, 1995
A>Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: Z20487; MIMD:95350813; PMID:7606788
A:Accession: T28625
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3006 <SU>
A:Cross-references: EMBL:L40609; NID:9886376; PID:9886377; PIDN:AAA75397.1
C:Gene(s):
A:Gene: var-3
A:Introns: 2597/3

Query Match	Score	771	DB 2:	Length	3006;
Best Local Similarity	41.58;	Pred. No.	3e-39;		
Matches 147;	Conservative	72;	Mismatches	115;	Indels 20;
					Gaps 6;
OY 1	YKEIQSYLSNDNKFVNINSEYKQFYKLTQVATNPFLNLMEGKCGKGLPGEK 60				
DB 392	YKRIEIKKSSSDKSNISNKNYENFGKKEVETLQVFLKLLKMGATCOKEIEBE 451				
OY 61	DIPTNSADKGIPTRESEYQVCPDCGVKCDGIKTHKSDNRER-----VNEDYKP 113				
DB 452	VIDF--NKDEDMVFRHREYCOPCDCVYCKGCKTEDKRNKCRSKTIKKIIQSEB--- 506				
OY 114	PMGKPTNITVLYSGNCGDITOKLENFCSSTYKRNKNOKWCYKDKENINCKLEON 173				
DB 507	-----PTEIHVNSDDKQGITKLEVFCSSTYBERNVOAKMKYKNSDYNCERNIS 561				
OY 174	TEINNDPKIISFNFPELWVTLRLPTIRKNDKLTCTINNFT--TTHCIDENRNCILCFDR 232				
DB 562	SYKSDTANVWLSEVCHSNKLNLLDITIKWEHQLKNCINNTVNTYCESKICKECYEK 621				
OY 233	WVQKKEEWSIKKLFTRKKNIQSYISNNLFEQGFYKMDLDE-AKKKELMENT 291				
DB 622	WIKRHEWEKVKVNFNNNMSYIYNNISRYVDSFLFQVMEFLDDDEKGMQOFIEDL 681				
OY 292	KRKKEESNLENNNDYLENAIELLDLKLKATICKDNTNEACETSHNATNP 345				
DB 682	K-KKEPSKTNTPYKGSQDAIEFLDLKDALTCRDNNSNESCVSKKYKTNP 734				

RESULT 2
T28161
hypothetical protein FC93-var11-1 - malaria parasite (Plasmodium falciparum) (fragment)
C/Species: Plasmodium falciparum

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C/Accession: T28157
R: Hernandez-Rivas, R.; Maltel, D.; Sterkers, Y.; Peterson, D.S.; Wellens, T.E.; Scherf, M.O. Cell. Biol. 17, 604-611, 1997
A>Title: Expressed var genes are found in Plasmodium falciparum subtelomeric regions.
A/Reference number: Z20483; MUID:97154495; PMID:9001213
A/Accession: T28161
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2647 <HER>
A/Cross-references: EMBL:U67959; NID:91794255; PID:91809295; PIDN:AA047438.1
A/Experimental source: strain FCQ27/PNG
C/Genetics:
A/Introns: 2158/3
A/Note: FCR3-varT11-1

Query Match 26.1%; Score 501.5; DB 2; Length 2647;
Best Local Similarity 34.4%; Pred. No. 8e-23;
Matches 116; Conservative 54; Mismatches 150; Indels 17; Gaps 11;

2 EKEIOSYLSNDKFNVNINSEYKQFEKLEKQYATNDPFLNLNKGKCGGLPGEKD 61
Db 395 ENELIRNNESSONSPKNNNNYETDFGNNKKDYQSMND-FLKLNSETPCTNIIDAKSK 453
QY 62 ITFTNSADDKGIFRSEYQVCPDQGVK--CDGKITYHKSDNDRERNNDYKPPMGVPR 119
Db 454 IDFTK--DPEFTSHTECDPCPCWGLKQADG-TWKRILENDPQCPKRYEPKGVPR 510
QY 120 TNLIVLSGNEOGDITQKLEFNCSSTNYDKNNQKWEYKDEINRCKLEQNTLEIND 179
Db 511 TEIDVLTGKENDIYKLEFCKTIDGNTGPK-NEEWNCYQVGN-DKCVLENGEELGGE 568
QY 180 NPKIISPHNFELWVYLLRDTIKMNDKLTCTINNTTHCIDCNRCICFEDWVAKKEE 239
Db 569 K-KVKDYDNFLMFWVAHMLKDSIEMRSKLSNCLSDKCTITTCNNCCQCYDKWIKKRY 627
QY 240 EWSIKKLFTRKKNIO-QSYYSNINLFEQ-YFEKVMKLDKDEAKWELMENIKRKNNE 297
Db 628 HWTQIKKHFKQDFOGNGRYFVLETVLEGDQFFDTITAYGDAEIVHIOEMLOKKKEQ 687
QY 298 FSNLENNRDYLENAIELLDLHLEKETATTC---KDN 330
Db 688 V-LHEDASNMKTIIDELDLHLEKAKQCIIVNKHNN 722

RESULT 3

Erythrocyte membrane protein 1 - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
A/Accession: T28157
R: Yang, S.Q.; Wollish, W.S.; Gut, J.; Wu, J.; Ahn, J.; Petersen, C.; Fujio, H.; Akawa
A/Description: The molecular cloning and DNA sequence analysis of Plasmodium falciparum
A/Reference number: Z20479
A/Accession: T28157
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2212 <YAN>
A/Cross-references: EMBL:U31083; NID:91517813; PID:91517814; PIDN:AA06961.1
C/Genetics:
A/Genes: EMP1

Query Match 26.0%; Score 499.5; DB 2; Length 2212;
Best Local Similarity 34.4%; Pred. No. 8.8e-23;
Matches 116; Conservative 56; Mismatches 148; Indels 17; Gaps 11;

QY 2 EKEIOSYLSNDKFNVNINSEYKQFEKLEKQYATNDPFLNLNKGKCGGLPGEKD 61
Db 395 ENELIRNNESSONSPKNNNNYETDFGNNKKDYQSMND-FLKLNSETPCTNIIDAKSK 453
QY 62 ITFTNSADDKGIFRSEYQVCPDQGVK--CDGKITYHKSDNDRERNNDYKPPMGVPR 119
Db 688 V-LHEDASNMKTIIDELDLHLEKAKQCIIVNKHNN 722

Db 454 IDFTK--DPEFTSHTECDPCPCWGLKQADG-TWKRILENDPQCPKRYEPKGVPR 510
QY 120 TNLIVLSGNEOGDITQKLEFNCSSTNYDKNNQKWEYKDEINRCKLEQNTLEIND 179
Db 511 TEIDVLTGKENDIYKLEFCKTIDGNTGPK-NEEWNCYQVGN-DKCVLENGEELGGE 568
QY 180 NPKIISPHNFELWVYLLRDTIKMNDKLTCTINNTTHCIDCNRCICFEDWVAKKEE 239
Db 569 K-KVKDYDNFLMFWVAHMLKDSIEMRSKLSNCLSDKCTITTCNNCCQCYDKWIKKRY 627
QY 240 EWSIKKLFTRKKNIO-QSYYSNINLFEQ-YFEKVMKLDKDEAKWELMENIKRKNNE 297
Db 628 HWTQIKKHFKQDFOGNGRYFVLETVLEGDQFFDTITAYGDAEIVHIOEMLOKKKEQ 687
QY 298 FSNLENNRDYLENAIELLDLHLEKETATTC---KDN 330
Db 688 V-LHEDASNMKTIIDELDLHLEKAKQCIIVNKHNN 722

RESULT 4

variant-specific surface protein 2 - malaria parasite (Plasmodium falciparum)
C/Species: Plasmodium falciparum
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
A/Accession: T28626
R: Su, X.Z.; Heatwole, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson,
A>Title: The large diverse gene family var encodes proteins involved in cytoadherence
A/Reference number: Z20487; MUID:95330813; PMID:7606788
A/Accession: T28626
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-2664 <SU>
A/Cross-references: EMBL:L40609; NID:9886376; PID:9886378; PIDN:AA75398.1
C/Genetics:
A/Introns: 2197/3
A/Note: var-2

Query Match 23.8%; Score 457; DB 2; Length 2664;
Best Local Similarity 31.1%; Pred. No. 4.3e-20;
Matches 114; Conservative 51; Mismatches 153; Indels 48; Gaps 9;

QY 2 EKEIOSYLSNDKFNVNINSEYKQFEKLEKQYATNDPFLNLNKGKCGGLPGEKD 53
Db 392 ENELIRNNESSONSPKNNNNYETDFGNNKKDYQSMND-FLKLNSETPCTNIIDAKSK 451
QY 54 -----GLPGEKDIPTFTNSADDKGIFRSEYQVCPDQGVKCDGKITYHKSDNDRERY 106
Db 452 IDFTKVOGGAIGGASDSNNNSKTFSSHQYCEPCGGVGLGNEM--KKNKGECK 509
QY 107 NNEIYKPPMGVPTNTIVLSGNEOGDITQKLEFNCSSTNYDKNNQKWEYKDEINR 166
Db 510 GKRRYINPDKHNVIPVLSFGDEHKEIKIDFOCAESMSDSKLTQOKCYGKEYE 569
QY 167 RCKLEQNTLEINDNPKT-ISPFPFELWVYLLRDTIKMNDKLTCTINNTT-HCIDCN 224
Db 570 VCTIENRNKSEEDPEEIOKTFHNFYFWIRHLLNDSIEMWKDNCIERAKGECKRNECK 629
QY 225 RNCIEDRWVAKKEEWSNKKLFTKKNIOQSYYSNINLFEQYFEKVMKLDKDEAKW 284
Db 630 TDCGCFQRIWIKKKEWGEIKKFK-----TQGFISFG--NYFVLENVLTID---- 677
QY 285 KEMLENIRKKNFSLNENRDYL-----ENAIELDLHLEKETATTCQDN 330
Db 678 -ELFQITTEYAGNSORIGIKDTLARKTQADDATEQKNTIDLLFEYSEAEKCK--K 734
QY 331 TNEACE 336
Db 735 IOBEQ 740

RESULT 5

T14029

[illegible]

681 CINNPKRKKCRNECKDDCGCFKEWIGRKKKEEMENIKKHFTQEAFFKNKRENSGIDMFSG- 739

```

681      | :| | | : :| | | :| | | : : |
      |||   | :| | | : :| | | :| | | : :| |
      CTUWDBZCBNACWDCCCEZTCZVZSTWUTZUZUSWACBDFEUCPCCHCCTDIAHCC - 730 .

```

```

        681 C I N N P K R K C R N E C K D D C G C F K E W I G K K K E E W E N I K K H F K T Q E A F K N K R E N S G I D M F S G - 739
        .

```

```
OY 270 FEFVMDKLD-----RDEKKWEIKMIRKKNESNL----- 301
      :|||          |         |||:: |||
Db 740 ---LMSADVLTELLEQLFQDINKGYDVAKELGINKIELDELKKKKGAEEAVVVVA 796
      :|||          |         |||:: |||
OY 302 ENNEDYLNNATIELLDLIKETAITCKDNNTNEACE 336
      :|: |||:: |||
Db 797 DNOK---RTTIDKLQHGGDANNCLTH--KECE 827
```

RESULT 8

Variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum) (fragment)
 N:Alternate names: erythrocyte membrane binding protein 1 (EMP1)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000
 C:Accession: J28155
 R:Rowe, J.A.; Moulds, J.M.; Newbold, C.I.; Miller, L.H.
 R:Nature 388, 292-295, 1997
 R>Title: Plasmodium falciparum rosetting is mediated by PfEMP1 and requires complement
 R:Reference number: J20477; PMID:97373957; PMID:9230440
 R:Accession: J28155
 A:Status: Preliminary; translated from GR/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2706 <ROW>
 A:Cross-references: EMBL:Y13402; PTDN:CAA73831.1
 A:Experimental source: strain IT 4/25/5
 C:Genetics:
 A:Introns: 2493/3
 A:Note: R29R+var1

Query Match	21.0%	Score 403.5	DB 2	Length 2706
Best Local Similarity	29.1%	Pred. No. 8.2e-17		
Matches 104; Conservative	69;	Mismatches 142;	Indels 43;	Gaps 14

[illegible]

RESULT 9

variant-specific surface protein 1 - malaria parasite (Plasmodium falciparum)
N:Alternate names: erythrocyte membrane binding protein 1 (EMPI)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T28432
R:Su, X.Z.; Heeltwale, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, D.S.
Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence and
A:Reference number: Z20487; MUID:95330813; PMID:7606788
A:Accession: T28432
A:Status: preliminary; translated from GB/EMBL/DBJ

A: Molecule type: DNA
A: Residues: 1-3078 <SUXY>
A: CROSs-references: EMBL: L40608; NID: g9886374; PTD: g9886375; PIDN: AAA75396.1
C: Genetics:
A: Gene: var-1
A: Introns: 2611/3

Query Match	20.6%;	Score 394.5;	DB 2;	Length 3078;
Best Local Similarity	26.2%;	Pred. No. 3.3e-16;		
Matches 122;	Conservative 76;	Mismatches 135;	Indels 133;	Gaps 20;

QY

```

Db      394 YDEEIKKXENGSAGSGSROKRDAGGTTTNNYDGYEKKFEYDELNKSEYRTVDKFELEKSMEE 453
QY      51 YC-----KGLPGKDLTFNNSA-----DDGIEFRSEYCOVCPDGCOKGDKXTH 97
Db      454 ICTKVKDEEGGIDTEKNNNSDSTSGASGNTVESOGTEFRSKCQCPPCYGVK---KVNN 509
QY      98 -KSDNDREYVNN-----EDKRPWGVKPTNTYVTSNNEGDDIQRKLENNC----- 143
Db      510 GSSSEWEKKNKNGCKSKGLYEPKPKDEGTTTTLKSKGHDDIEKLNKFCODEKNGDTI 569
QY      144 ----SSTVWKDKNN-----QKMECYKDEINRCKLEONTIENNDPK-----II-- 184
Db      570 NSGGSTGGSGGNGNSRQRELYHEWKC-YKGEYVAVGVHDEDEFEYENVKVNAAGCILIIC 628
QY      185 -----SFHNFELWYTYLLRDYIKMNDKLTCTI-NNTTTHC-ID 221
Db      629 QKKNKEBEGNISEKEPDELQKTENPFYFYVAHMLKDSIHMKKTKQROCLONGNRIRKCGNN 688
QY      222 EGNRNCCLPDRVWVOKKEEWNISIKKLTFFKKNNIQ-----QSYISNNLNF 267
Db      689 KKNNDCECFKRWITQKKDEWCKIYQHF-KTONIKRGSGSDMTALPIPDHDYVLYADQIE 747
QY      268 GYF-----FKYMDLIDQAKKMLMEIRKKKKKNEFSLENNRD-----YLEN 310
Db      748 EFLKGDSEDAESEKSENSIDAEEA-----EELKHLREIIBESDENNOBSVGGVTEQKN 801
QY      311 ATELLIDHLKETATIC-----KDNNTNACEISHNATNP 345
Db      802 IMDKLINTYEKADLCLEIHEDDEEKEKGGDNECIEBENFRYP 847

```

RESULT 10

variant-specific surface protein 7 - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: J28634
R:Su, X.Z.; Heitwohle, V.M.; Wertheimer, S.P.; Guinet, F.; Herrfeldt, J.A.; Peterson, Cell 82, 89-100, 1995
A:Title: The large diverse gene family var encodes proteins involved in cytoadherence
A:Reference number: J20487; MUID:95330813; PMID:7606788
A:Accession: J28634
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Residues: 1-2182 <SU>
A:Molecule type: DNA
A:Cross-references: EMBL:J28636; NTD:g886379; PID:g886380; PIDN:AAV5399.1
C:Genetics:
A:Note: var-7

Query Match	19.8%	Score	380;	DB 2;	length	2182;			
Best Local Similarity	28.7%	Pred	No. 1.8e-15;						
Matches	109;	Conservative	55;	Mismatches	148;	Indels	68;	Gaps	12

```
OY      1 YKKEL-----QSYLSDNKFVNINSEYYKCEYEKKEQVATNTDF 43
         ||| | | | | | | | | | | | | | | | | | | | | |
Db      413 YTTETISGGSGSKPKRTKRAARSSSSDDN-----GYESKFEYKKLKEVGyGDVDVKFL 466

OY      44 NILNGSKTCYG---GLFGEKDITFTNSADNDKGIFRSEYOQVPRDCCYAKDKGIKITYHKS 99
         ||| | | | | | | | | | | | | | | | | | | | | |
Db      465 KILNEGICIQKQPOVGNENKAADVDTNNKIYK-FTSRTEIDPCWCGLEGGGPPRWYK 523
```

QY 100 DNDREANNEDYPPWGVKPTNTVLYSGNEGDIYOKLENFCSSTNYKKNQKWEY 159
 Db 524 DKTGSAKTYTPK---NTTDPVLYPKSOONILTKKYNFCEKGNP--GGGIRKQWY 579
 QY 160 YRD-----ENINCKLEONTNNDNPRTISPNFELWTVLYLRPTIKMDLKTICNN 214
 Db 580 YDEHRSSKNNNNCCVEGTWDFWOGKQYKSYNFWFMDVHMDLHDSVEKTELKSCINN 639
 QY 215 TTT---HCIDECNRNCLCFDRWVYKQKEEWMNISKLTFTKKNIOQSYSNINNEGYF 270
 Db 640 NTNGNCRNNKCKTDCGCGQKQWKEKKQOEWMIAIKHFGKQGTIYQO-----KGLYFSP 694
 QY 271 FKYMDLKDDEAKWKELEMIKRRKNEFSLNENRDTL-----NALELL 316
 Db 695 YGVLDLY---LKGMLQWIKQVHGTDDIKIKILLDEDAVAAYVAGSKDWTIDKL 750
 QY 317 DHEKETATICKDNNTNFACE 336
 Db 751 QHEKEQAGCKOK--QEECE 768

RESULT 11

variant-specific surface protein 1 homolog PFB0010w - malaria parasite (Plasmodium falci-
 C:Species: Plasmodium falciparum
 C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C:Accession: C71625
 R:Gardner, M.J.; Tetteh, H.; Garuol, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 Science 282, 1126-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; M0ID:99021743; PMID:9804551
 A:Accession: C71625
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-1711 <GAP>
 A:Cross-references: GB:AE001366; GB:AE001362; NID:g3845070; PIDN:AACT1792.1; PID:g384507
 A:Experimental source: clone 3D7
 C:Genetics:
 A:Gene: PFB0010w

Query Match 18.9%; Score 363; DB 2; Length 1711;
 Best Local Similarity 27.5%; Pred. No. 1.5e-14;
 Matches 100; Conservative 60; Mismatches 130; Indels 74; Gaps 14;
 QY 1 YEKEIOSYLSNDKFNVINSEYKQFEKLEKTOYATNDTFLNLNEGKCGGLP-GE 59
 Db 424 YEHFNEELNEGKDR-----SFLQLSKERICKERIOVGE 460
 QY 60 KDITFTNSADKGIYRSEYQVCPDGVKCDGIKYTHKSDND--REVRNNEYKPPWGVK 118
 Db 461 ETANYGNFENESMTFSITECDRCPLCGVDCSSDNCRRKDKSGDEIDTKEYPE--N 517
 QY 119 PTNITVLYSGNEGDIYOKLENFCSSTNYKKN---QKWEY---KDE---NINR 167
 Db 518 TTRIPKLTAKRRTGLIKKEKFKNS---DSNGQIKKWECHYKKNKDKDNGDINN 573
 QY 168 C-KLEQTEINNDNPRTISPNFELWTVLYLRPTIKMDLKTICNNTT-HCIDECNR 225
 Db 574 CIGODKMTSNVYR--ISYSPFYGITIDMLNLSIEMRELSICINDALGCKRCKCN 631
 QY 226 NCICFDWVQKEEWMNISKLTFTKKNIOQSYSN--NINLFEYFPPVMDKLDKDAKW 284
 Db 632 PCCEYKRWVKKKDEMDKIKIEFFRKQDLKLAGMDAGLEYLENI-----680
 QY 285 KELEMIKRRKNEFSLNENRDTLENAIELLDHLKETATL-----CKDNT 331
 Db 681 --FLEDMKRNAGDPKYLEKEFKLENGEYODPLKTKTKTIDFLKELENAKCKVEKNP 738
 QY 332 NEAC 335
 Db 739 DNEC 742

RESULT 12

T14602

variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragment
 C:Species: Plasmodium falciparum
 C>Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jun-2000
 C:Accession: T14602
 R:Yoss, T.S.; Felger, I.; Wells, N.; Beck, H.P.
 submitted to the EMBL Data Library, February 1998
 A:Description: Identification of a conserved 5' flanking region of Plasmodium falcipa-
 A:Reference number: Z18158
 A:Accession: T14602
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2135 <VOS>
 A:Cross-references: EMBL:AF050740; NID:g2944094; PID:g2944095; PIDN:AA05220.1
 C:Genetics:
 A:Gene: varph17

Query Match 18.8%; Score 361.5; DB 2; Length 2135;
 Best Local Similarity 27.5%; Pred. No. 2.4e-14;
 Matches 111; Conservative 57; Mismatches 142; Indels 93; Gaps 16;

QY 1 YEKEIOSYLSNDKFNVINSEY---KQFEKLEKTOYATNDTFLNLNEGKYC--K 53
 Db 401 YGEI-SLKSRRKRDAGISTKYVDGEKFEYELK-SEYRVGEFLGLNNEKTCREK 458
 QY 54 GGLPEGDIYFTNSADK-----IFYRSEYQVCPDGVKCDGIKYTHKSDNDREYNN 108
 Db 459 DKEGKIDFTVNSGSASGDVKNFTFRYCECPGCAEOERNGVWAKAKDRDCSPG 518
 QY 109 EDKPPWGVKPTNTVLYSGNEGDIYOKLENF-----142
 Db 519 NDYTK---YKKKEIPILTGTDKTSEIYERKRCCKNNKNGANGREGVSGEAGSANS 575
 QY 143 -NSTNYKDKNN-----QKWEY---KDEN-----INCKLEONTNNDNPRTISFN 188
 Db 576 DNATGTGCGGNSDSSLCSEKFWTCYKRRKENNKGNDINFC-VOGAMQNSKDKQKYSNA 634
 QY 189 FPELWTVLYLRPTIKMDLKTICN--NTTHCIDECNRNCLCFDRWVYKQKEEWMNISK 246
 Db 635 FFWDMVHMDLIDSKWNEHGKCIKNDNGKT-CIKGCKSKDCGLKVVQKEKEMKLTLE 693
 QY 247 LFTKKNIOQSYSNINLFEYFPPVMDKLDKDAKWELEMIKRRKNEFSLNENRD 306
 Db 694 HFTMGGEFGDGEHQRGFTHDVYLVNLLDK-----KELKTI---EGTGTETETKH 742
 QY 307 YLE-----NALELLDHLKETATCKD 328
 Db 743 IKELDKREERDAGTGVAAGTGPRNIMDKLIEHQLQAKKCKD 785

RESULT 13

T18396

erythrocyte membrane protein variant 2 - malaria parasite (Plasmodium falciparum) (fr
 C:Species: Plasmodium falciparum
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18396
 R:Baruch, D.I.; Pasloske, B.L.; Singh, H.B.; Bi, X.; Ma, X.C.; Feldman, M.; Taraschl,
 Cell 82, 77-87, 1995
 A:Title: Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen a
 A:Reference number: Z18925; M0ID:95330812; PMID:7541722
 A:Accession: T18396
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1729 <BAR>
 A:Cross-references: EMBL:U027339; NID:g914920; PID:g914921; PIDN:AAA89134.1
 C:Genetics:
 A:Gene: EMP1
 A:Note: var-2

Query Match 18.6%; Score 357.5; DB 2; Length 1729;

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:46 ; Search time 3.31492 Seconds

(without alignments)
4316.649 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1918
Sequence: 1 YEKIQSYLSNDNKFVNNIN.....CKDNTNACETSHNATNP 345

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

a1 number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	7.8	1005	1	RASO_METUA
2	146	7.6	2339	1	RPCI_PLARA
3	140	7.3	2136	1	YCR2_MARPO
4	137	7.1	1435	1	EBAL_PLAFC
5	136	7.1	1147	1	CGAL_HELPY
6	130	6.8	1956	1	ATXI_HELPY
7	127.5	6.6	1790	1	USOI_YEAST
8	122	6.4	1805	1	HMW2_MYCGE
9	121	6.3	1153	1	PVDB_PLAKN
10	120.5	6.3	537	1	ARP_PLAFA
11	119.5	6.2	782	1	Y091_CAEBL
12	119	6.2	971	1	Y228_BORBU
13	119	6.2	1073	1	PVDA_PLAKN
14	118.5	6.2	1085	1	IFPI_YEAST
15	118.5	6.2	1220	1	IFZE_HUMAN
16	118	6.2	1070	1	PVDR_PLAVS
17	117	6.1	500	1	RT03_PROWI
18	116	6.0	1130	1	YL17_CAEBL
19	115.5	6.0	864	1	RASO_SULSO
20	115	6.0	494	1	PRE_STRAG
21	114.5	6.0	1202	1	RPW2_STRELOC
22	114.5	6.0	1280	1	KCPG_XENLA
23	114	5.9	1701	1	MSPI_PLAFM
24	113.5	5.9	537	1	Z029_XENLA
25	113.5	5.9	816	1	YG3A_YEAST
26	113	5.9	357	1	TRMA_CAMJE
27	113	5.9	766	1	STRB_YEAST
28	113	5.9	1186	1	CAGA_HELPY
29	112	5.8	709	1	CDAT_PLAFA
30	112	5.8	1036	1	Y414_MYCGE
31	112	5.8	1726	1	MSPI_PLAFA
32	111.5	5.8	964	1	DPOL_CBEPV
33	111.5	5.8	1082	1	SP23_YEAST

34	111.5	5.8	1251	1	RB2_PLAVB	000799 plasmodium
35	111	5.8	1038	1	CIN8_YEAST	P27895 saccharomyc
36	111	5.8	1726	1	MSPI_PLAFC	P04934 plasmodium
37	110.5	5.8	1070	1	PVDC_PLAKN	P50494 plasmodium
38	110.5	5.8	1164	1	BAG_STRAG	P27951 streptococc
39	110	5.7	1169	1	EXSB_BORBU	051378 borrelia bu
40	109.5	5.7	609	1	YSW1_YEAST	P38280 saccharomyc
41	109.5	5.7	630	1	Y242_MYCGE	P47484 mycoplasma
42	109	5.7	886	1	RASO_SULAC	033600 sulfolobus
43	108.5	5.7	731	1	BAFI_YEAST	P14164 saccharomyc
44	108	5.6	451	1	ARP2_PLAFA	P13824 plasmodium
45	108	5.6	1076	1	RPOB_ASTLO	P27059 astasia lon

ALIGNMENTS

RESULT 1
RASO_METUA STANDARD; PRT; 1005 AA.
AC 058718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA double-strand break repair rad50 AtPase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschli.
OC Archaea: Euryarchaeota: Methanococci: Methanococcales;
OC Methanocaldococcaceae: Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.L.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Usterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.,
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
jannaschli.";
RT Science 273:1058-1073(1996).
RL -1- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
rad50/mre11 complex possesses single-strand endonuclease activity
and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
and/or repositioning DNA ends into the mre11 active site (By
similarity).
CC -1- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -1- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: U67572; AAB99331.1; .
CC TRIGR; MJ1322;
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02463; SMC_N; 1.
DR Pfam: PF02483; SMC_C; 1.
DR ProDom: PD000006; ABC_transportr; 1.
DR DNA_repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (By similarity).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).

SQ SEQUENCE 1005 AA; 119367 MW; 9BBDB48173E788F3 CRC64;
 Query Match 7.8%; Score 149; DB 1; Length 1005;
 Best Local Similarity 21.8%; Pred. No. 0.021;
 Matches 93; Conservative 53; Mismatches 109; Indels 172; Gaps 19;
 QY 13 NKPNNINSEY--YKQYEEKIKETQYATNDFTLNLNEGKYGKGLGPEKIDITFTNSAD 70
 DB 475 NKKIGETINSEIKRLKILDELEK-----EGK----- 501
 QY 71 KGIYRESEYCOVPCDGVKCGIGITVTHKSDNDRRNVNEDYKPPWGVKPTNITYVLSGNE 130
 DB 502 -----CPLCTPTIDENKMKMLINQHKTKQLNK-----YEELEIRNK 537
 QY 131 -----OGDITQ-----KLENFCSSTNYKDKNNQKWECCYKKNRCKLEQNTN-- 177
 DB 538 KIRIEIKDIKKEIKEDIKENLKTLLYLEKQSIELEIKLKNYK-----EQDELIRNK 533
 QY 178 -----NDNP-----KIISFHFELM--VYTL-----LRDTIK----- 203
 DB 594 ISNIVYNGKFPVDELIEDIKSQLNKFKNFYMQYLSAVSYNSVDEGIRNRKIEINIVSG 653
 QY 204 WN-----DKLKTGIN--NTTHCIDECNRCICFDRW-----VKOK 237
 DB 654 WNKKECREELNKLREDEREINRLKDKLNEKNEKELEIENRSLKFDKYKEYGLITER 713
 QY 238 EEEWNSIKKLFYTKKNIOOS-----YYSN-----INLFE--G 268
 DB 714 LELKLNKIKDGLIEIYVNCNSKILAIIDIKRKYKEDIEIYLNKILEVKNKEINDIERIS 773
 QY 269 YEFKVDKIDKDAKKELENIENIKRKKNEFSN-----IENRDLLENAILLIDHKE 321
 DB 774 YINOKIDEIYNNEEKKIKELIETENKROELDNVBEQTELEIGIYELKDKVESLKARKE 833
 QY 322 TATICND 328
 DB 834 MSNLEKE 840
 RESULT 2
 RP1_PLAFA STANDARD; PRT; 2339 AA.
 ID RP1_PLAFA STANDARD; PRT; 2339 AA.
 AC P27625;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNA-directed RNA polymerase III largest subunit (EC 2.7.7.6).
 OS Plasmodium falciparum.
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5833;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92018020; PubMed=1656254;
 RA Li W.B., Bazik D.J., Tanaka M., Gu H., Fox B.A., Inselburg J.;
 RT "Characterization of the gene encoding the largest subunit of
 RT Plasmodium falciparum RNA polymerase III.";
 RL Mol. Biochem. Parasitol. 46:229-240(1991).
 CC -1- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
 CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
 CC SUBSTRATES.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate - N diphosphate +
 CC (RNA)(N).
 CC -1- SUBUNIT: RNA POLYMERASE III CONSISTS OF ABOUT 15 DIFFERENT
 CC SUBUNITS. THIS SUBUNIT IS THE LARGEST COMPONENT OF RNA POLYMERASE
 CC III.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES ARE
 CC FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
 CC PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
 CC III FOR 5S AND TRNA GENES.
 CC -1- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M73770; AAA29729.1; -
 DR InterPro; IPR000722; RNA_pol_A.
 DR InterPro; IPR002879; RNA_pol_A2.
 DR Pfam; PF00623; RNA_pol_A; 1.
 DR Pfam; PF01854; RNA_pol_A2; 2.
 KW Transferase: DNA-directed RNA polymerase; Transcription; zinc;
 KM Zinc-finger; Nuclear protein.
 FT ZN.FING 88 101
 SO SEQUENCE 2339 AA; 272829 MW; EDD899363086FD48 CRC64;
 Query Match 7.6%; Score 146; DB 1; Length 2339;
 Best Local Similarity 20.2%; Pred. No. 0.079;
 Matches 79; Conservative 63; Mismatches 128; Indels 122; Gaps 20;
 QY 9 LSNDRKFEVNNINSEYKQ--FYEKL-----KETQYATNDFTLNLNEGKYGKGLPGE 59
 DB 1229 LEHDEFLVKNKIMESYTDQCNEDYEDIRSLDKNYSYIHD-----QKHLSLQCAE 1281
 QY 60 KIDFTNSADDKGIFYRSEYCOVPCDGVKCGIGITVTHKSDNDRRNVNEDYKPPWGVK 119
 DB 1282 EHTITNTNNDN-----TYEQLIEMELSKNKTREKQSPFG 1317
 QY 120 TITVYVLSGNEGD--ITQKLENF-----CNSSTNYKDKNNQKWE-----C- 158
 DB 1318 TYRDMHEDESEOMNFEITRKAFIEKKKGKMHCCNDIEY--NNQYDNIQYNNISCN 1374
 QY 159 YKDEININCKLEONT--EINND--NPKII--SHNFPELMWYLLADDTIKW--- 205
 DB 1375 YIKSONL-----ENTHQVNDLSFIKNVILPREKHSIFHFVNDY--RNVEIKKIM 1426
 QY 206 DKLKTCINNTTHCIDECNRCICFDRWVKQKEEWSNIRKLFYTKKNIOQSYNSINNL 265
 DB 1427 DKKRIFLNNSEKNV-----QSYNNSKSNLKKIKELI-----INNT 1462
 QY 266 EGYFFKY--MDKIDKDEAKY-----KELMENIKRKKNEFSNLENNRDYLENAIE 313
 DB 1463 YRNEKKKILNRKTKKNDNDNTWSSDDSIIAKKIKIKERKRYHPRKEKENEDRNRYK 1522
 QY 314 LILD-HLKETATICKDNTNTACETSHNATIN 344
 DB 1523 MITDNNNNNDNNNNNDNNNNNDNNNNNNNNNN 1554
 RESULT 3
 YCF2_MARPO STANDARD; PRT; 2136 AA.
 ID YCF2_MARPO STANDARD; PRT; 2136 AA.
 AC P09975;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 259 kDa protein ycf2 (ORF 2136).
 GN YCF2.
 OS Marchantia polymorpha (Liverwort).
 OG Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Marchantiophyta;
 OC Marchantiopsida; Marchantiales; Marchantiaceae; Marchantiaceae;
 OC Marchantiaceae; Marchantia.
 OX NCBI_TaxID=3197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Ohyanaka K., Fukuzawa H., Kohchi T., Shirai H., Sano T., Sano S.,
 RA Umesono K., Shiki Y., Takeuchi M., Chang Z., Aota S., Inokuchi H.,
 RA Ozeki H.;
 RT "Chloroplast gene organization deduced from complete sequence of
 RT liverwort Marchantia polymorpha chloroplast DNA.";
 RL Nature 322:572-574(1986).

```

CC -1- SIMILARITY: BELONGS TO THE YCF2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X04465; CAA28078.1; -
DR PIR: A05037; A05037.
DR PIR: S01591; S01591.
DR InterPro: IPR003959; AAA_ATPase_cent.
DR Pfam: PF00004; AAA: 1.
KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 2136 AA; 259911 MW; 5BD170C9CCF61197 CRC64;

Query Match
Best Local Similarity 7.3%; Score 140; DB 1; Length 2136;
Matches 92; Conservative 51; Mismatches 91; Indels 160; Gaps 24;

QY 4 EIOSYLSNDKRFY-----NINSEYKQFY-EKLEK-----OYATN 39
DB 508 EILSINEIDKPFYINKISLKNKKOKRYLKNKSSDNFRPLMKIKNYSSQCFVSN 567
QY 40 DTEL-----MLNEGKCKGGLPGEDITFTNSAD--KGIFYRSEYQVCPDGVKCDG 92
DB 568 NSFLNDPAFEILQONYLK-----KNILFCKLNEVFSNFFGYOY-----KC-- 611
QY 93 IKTHSHSDNDRVNNEDYKPPMGVKEPTN-----TVYSGNDOGITOKL-ENFCN 143
DB 612 -----KLNIFLKFASLEKILKRNKFTLSIKLKFKE-- 644
QY 144 SSTYKDKNNQKCEYKIDEN--INCKLEONTEN--NDPKITSPNFELWATYL 197
DB 645 -----YKNKLNENGE--YKISQLONEKEKLNKKRKNFOFNPKILSFTNSK--KNYI 696
QY 198 LRDITKMDLKTCTINNTTHGIDECNRNCLCFDRWYKKEE-----ENNSIKLEPTK 251
DB 697 LQNKYEFN--KNLNNKLI-----TWKIKSNKLVIENSEYN--KIITWK 736
QY 252 KNIQOASYNNINLKESEYFK-----VMDKLDKDEAKKELM-----EN 290
DB 737 KNNKFESEK--NSVLDTEFFNKKSPNITVIEFKLKIQLNFOELKILNCFSLFNSKN 795
QY 291 IKRRK-----NE-----FSNLNNRDYLE 309
DB 796 IKTKIKFKNSYFINENITTFESFNDRKFNIFLE 829

RESULT 4
EBAL_PLAFC
ID EBAL_PLAFC STANDARD; PRT; 1435 AA.
AC P19214;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Erythrocyte-binding antigen EBA-175.
OS Plasmodium falciparum (isolate Camp / Malaysia).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5835;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90377299; PubMed=2204835;
RA Sm B. K. L.;
RT "Sequence conservation of a functional domain of erythrocyte binding
RT antigen 175 in Plasmodium falciparum."
RL Mol. Biochem. Parasitol. 41:293-296(1990).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X52524; CAA36756.1; -
DR PIR: S11561; S11561.
DR Antigen.
FT DOMAIN 159 1104
FT ERYTHROCYTES.
FT VARIANT 1031 1031 E -> V (IN STRAINS PCR-3 AND ITG).
SQ SEQUENCE 1435 AA; 167389 MW; 32A4309021B1C3D6 CRC64;

Query Match
Best Local Similarity 7.1%; Score 137; DB 1; Length 1435;
Matches 86; Conservative 57; Mismatches 149; Indels 128; Gaps 19;

QY 1 YEKIQTSLSDN-----KRVN-NINSEYKQFY-EKLEKQYATNDPFLMLNEGKCYK- 53
DB 75 FNKRWKSYGTPDNDKMSLKNKNNEMFNNNYQSLSTS-----SLIKONKQVPI 126
QY 54 GLPGEKDIPTFTNSADDKGIFYRSEYQVCPDGVKCDGIKYTHKSDNDR-VNEDYK 112
DB 127 NAYAVSHILSPDSRINGG-RNTSSNNEVLSCNEKKRGMKMDCKKKNDSNIVCIPDR 185
QY 113 PPMGVKPTNITVLYSGNEQGD-----ITQKLEN-----FCNSSTN----- 147
DB 186 IQCLVMLSLIKTYTKFTMDHFEASKESQLLKKNNDKNYKFCNDLKNSLDGYHL 245
QY 148 -----YDKNNOK-----WECYKQD--E 163
DB 246 AMGNMDQEGYSTKAENKIOEVPFAGAGEISEHKIKNFKEMWNEPREKLEAMLSHK 305
QY 164 NINCKLEONTENINDNPKIISFNPELWATYLRDITKMDLKTCTINNTTH-CIDE 222
DB 306 NINCKNIPQEL-QITOWIKEMHGER-----LIERNRSKLPSK--CKNNTLYEACEKE 358
QY 223 CNRNCLEFDRWYKQKEEWSIKLFTKKNKNIQOASYNNINLPEGEYFVMDRLDDEA 282
DB 359 CIDCMKRPWIIISKEFWHTLSKEYEFOKVPKEN-----AENYLIKISE--NKDA 408
QY 283 KWKLEMINIRKKNKEPNNLENNRDYLENAIYELLDLHLEKATICK-----DNTNACE 336
DB 409 KVSLLNNDQAEYSKYD-----CKHTTIVKSVLNGNDNTIKERKE 450

RESULT 5
CGAL_HELPY
ID CGAL_HELPY STANDARD; PRT; 1147 AA.
AC P80200;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Cytotoxicity associated immunodominant antigen (120 kDa protein).
GN CAGA OR CAT.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CCUG 17874 / NCTC 11638;
RX MEDLINE=93296225; PubMed=8516329;
RA Covacci A., Censini S., Bugnoli M., Petracca R., Burroni D.,
RA Macchia G., Massone A., Papini E., Xiang Z., Figura N.,
RA Rappunoli R.;
RT "Molecular characterization of the 128-kDa immunodominant antigen of
RT Helicobacter pylori associated with cytotoxicity and duodenal
RT ulcer."
RL Proc. Natl. Acad. Sci. U.S.A. 90:5791-5795(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CCUG 17874 / NCTC 11638;

```

RX MEDLINE=97121442; PubMed=8962108;
 RA Censini S., Lange C., Xiang Z., Crabtree J., Ghiara P.,
 RA Borodovsky M., Rappuoli R., Covacci A.;
 RT "cag, a pathogenicity island of *Helicobacter pylori*, encodes type I-
 RT specific and disease-associated virulence factors.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:14648-14653(1996).
 RN [3]
 RP SEQUENCE OF 181-190; 314-328; 366-377; 420-430; 494-506; 661-677;
 RP 900-914 AND 1062-1077.
 RA Hermann V., Hermann J., Kist M.;
 RL Submitted (APR-1993) to the SWISS-PROT data bank.
 CC -I- FUNCTION: MAY BE NECESSARY FOR THE TRANSCRIPTION, FOLDING, EXPORT,
 CC OR FUNCTION OF THE CYTOTOXIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).

DR	EMBL; X/0039; CAA49633.1; -.				
DR	EMBL; AF282853; AAC44706.1; -.				
DR	InterPro; IPR005169; Caga.				
DR	InterPro; IPR004355; IVSec_caga.				
DR	Pfam; PF03507; Caga; 1.				
DR	PRINTS; PR01553; TYPE4SSCAGA.				
KW	Antigen.				
FT	DOMAIN	880	885	POLY-ASN.	
FT	CONFLICT	320	320	G -> A (IN REF. 3).	
FT	CONFLICT	325	325	R -> F (IN REF. 3).	
FT	CONFLICT	328	328	P -> K (IN REF. 3).	
FT	CONFLICT	426	426	K -> E (IN REF. 3).	
FT	CONFLICT	429	429	Q -> E (IN REF. 3).	
FT	CONFLICT	673	675	AOX -> TED (IN REF. 3).	
FT	CONFLICT	901	901	A -> T (IN REF. 3).	
FT	CONFLICT	903	903	Q -> E (IN REF. 3).	
FT	CONFLICT	907	907	L -> P (IN REF. 3).	
FT	CONFLICT	910	910	P -> R (IN REF. 3).	
FT	CONFLICT	914	914	Q -> E (IN REF. 3).	
FT	CONFLICT	1072	1072	P -> S (IN REF. 3).	
FT	CONFLICT	1074	1074	S -> D (IN REF. 3).	
SO	SEQUENCE	1147 AA;	128013 MW;	AB92770835F68490 CRC64;	

Query Match	7.1%;	Score 136;	DB 1;	Length 1147;
Best Local Similarity	19.0%;	Pred. No. 0.16;		
Matches	84;	Conservative	69;	Mismatches 139;
				Indels 150;
				Gaps 17

```

2 EKEIOS-----YISNDKFEVNNINSEYKOFYKLE-----33
Db 390 QOEIÖNK IDMEPLAÖNNAK IDNISEKBEKEFTEIKDOFOKDSKAYI DALGNDRIAPVSK 44.9
QY 34 -----POYAINDTFLNLLNGKCYKGGGLPBEXKOTIFITNSADMDGIFITRSEYQYCP 84
Db 450 KDTYHSA LITIEGMDGSLYTLKIDYGRADKADALDEKREKNTLOGSLKHGVMF-----500
QY 85 DCGVCKDGIKTYHRKSDNDRE-----RYNNE 109
Db 501 ---DYSNFEKTYNASKNPNKRGVGTNGVSHLEVGFNKVAIFENLPLDNLNLAITSFVRNLE 557
QY 110 DYKFPWGKPKENITIVL-----YSGEGÖDIIQKLENF-----CHSSYNYK-----D 150
Db 558 DKLITTKGSLPEANKLIKDFLSSNK---ELVGKTLNPFKNAVADAANKNTNYDEVRKKAÖDLE 615
QY 151 KNNOKWECYVAD-----ENINCKLEQÖNTEINNDPKIISFHNFELWYTYLLRD--T 201
Db 616 KSLRREHLEKEVEKLELESKGNKNKMEAKQAÖNSÖDEIFALIN-----KEANRDARA 666
QY 202 IKAMDKLTCTCINNNTTHCIDCÖNCCLCFDPMVQÖKEEÖNNSIKKLEFTKKKNNOÖSYSN 261
Db 670 IAYÖNÖLNG-IKRELIDSKLEVANNLNLDFF---KSPEFNKGKNDKSKAEETIKALGS 723

```

[illegible]

	RESULT 6
ATX1_PLAFA	
ID	ATX1_PLAFA
AC	004956
DT	30-MAR-2000 (Rel. 39, Created)
DT	30-MAY-2000 (Rel. 39, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Probable cation-transporting ATPase 1 (EC 3.6.3.-).
OS	Plasmodium falciparum.
CC	Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
OX	NCBI_TaxID=5633;
EN	[1] Unpublished report

RP SEQUENCE FROM N.A.
RC STRAIN-T9/96;
RX MEDLINE=93132070; Pubmed=8421054;
RA Krishna S., Cowan G., Meade J.C., Wells R.A., Stringer J.R.,
RA Robson K.J.;
RT "A family of cation ATPase-like molecules from *Plasmodium falciparum*";
RT J. Cell Biol. 120:385-398(1993).
RL -1. CATALYTIC ACTIVITY: ATP + H(2)O = ADP + phosphate.
CC -1. SUBCELLULAR LOCATION: Integral membrane protein.
CC -1. SIMILARITY: BELONGS TO THE CATION TRANSPORT ATPASES FAMILY (E1-E3 ATPASES). SUBFAMILY V.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed, usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL: X65738; CAA46646.1; -.
DR InterPro: IPR001757; ATPase_E1-E2.
DR Pfam: PF00122; E1-E2_ATPase; 1.
DR PROSITE: PS00154; ATPase_E1_E2; 1.
KM Hydroxylase; Transmembrane; Phosphorylation; Magnesium; ATP-binding
Protein; 1 25

FT	DOMAIN	1	35	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	36	58	POTENTIAL.
FT	DOMAIN	59	61	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	62	80	POTENTIAL.
FT	DOMAIN	81	407	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	408	427	POTENTIAL.
FT	DOMAIN	428	440	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	441	462	POTENTIAL.
FT	DOMAIN	463	1818	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1819	1837	POTENTIAL.
FT	DOMAIN	1838	1845	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1846	1863	POTENTIAL.
FT	DOMAIN	1864	1881	CYTOPLASMIC (POTENTIAL).
FT	TRANSSEM	1882	1905	POTENTIAL.
FT	DOMAIN	1906	1928	EXTRACELLULAR (POTENTIAL).
FT	TRANSSEM	1929	1952	POTENTIAL.
FT	DOMAIN	1953	1956	CYTOPLASMIC (POTENTIAL).
FT	MOD. RES.	496	496	PHOSPHORYLATION (BY SIMILARITY).
FT	METAL	1760	1760	MAGNESIUM (BY SIMILARITY).
FT	METAL	1764	1764	MAGNESIUM (BY SIMILARITY).
FT	DOMAIN	246	251	POLY-ASN.
FT	DOMAIN	252	256	POLY-LYS.
FT	DOMAIN	937	941	POLY-ASN.
FT	DOMAIN	1344	1347	POLY-LYS.
FT	DOMAIN	1363	1372	POLY-ASN.
FT	DOMAIN	1680	1684	POLY-ASN.

SEQUENCE 1956 AA; 230285 MW; AE708AAE9009335 CRC64;
 Query Match 6.8%; Score 130; DB 1; Length 1956;
 Best Local Similarity 19.2%; Pred. No. 0.68;
 Matches 90; Conservative 54; Mismatches 150; Indels 174; Gaps 19;

1 YE-KEIGSLND-----NKFVNNINSEYKQYKLETOYATNDFLNLNBSKYK 53
 1273 YEERKNIYVHIMKKKKIKNIINNNSNLY-----YHNITIDTFVKRMKREYCF 1323
 54 GGLPEKEDIFETNSADKGI-----YREXCQYCPDGVKCDGKITYHKSDND 102
 1324 NKL-----LYKIOCKLLNLIHLNLYKKKKYNNNYDI-----DEVHLIGNNNN 1367
 103 RERVNEDYRPMGVKPTNITVLYSGNEGD-----ITCKLEFNCN 143
 1368 NKNNSKCKPLKKNKKH-----RKNESNDNTFNTYSNNIHLISKYVYHKNYYPD 1423
 144 SSTRYKDKNN-----OKWECYKDEININCKLEONTNINNDPKT-----ISF 186
 1424 SCNLRKKKNSLFYKLYYERKYLQCLKHDKYKVELPRIDINYSYQMESIKT 1483
 187 HFEPE-----LWVYLLR--DTIKMNDKL--KTCINNTTHCIDECNRNCLCF 230
 1484 RNFHLSSEQPAFNSNLLSYIIKNDNNVYNNKYNNKYNNKNS-----ICKNATCN 1539
 231 DRWYKOREEWN-----SIRKLF-----TKKKN 254
 1540 KNYIYNNNNIYNNKNIYNNKNIILHAKSVLLSGSKKFKFSNIIIRHKLKKNKNI 1599
 255 QOQSYNNINLFGYFPRKVA-----DKIDKDAKKEKLEMEIKR----- 293
 1600 KRKMKNHVNNTSGHILLNCTHGFKDYSSLNKRYIVNNKRYMLNDNYDRHMYLT 1659
 294 -----KKNFESNLENNRDYLENAIELLDHLEKETAATCKDN 329
 1660 DWYRGTOYGSKKKKNKIYNNNNNNILKKNINRLEHL--LVDKCKRN 1705

RESULT 7
 USOL YEAST STANDARD; PRT; 1790 AA.
 ID USOL YEAST
 AC P25386;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Intracellular protein transport protein USOL.
 EN USOL OR INT1 OR YDL058W.
 CC Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CC NCBI_TaxID=4932;
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=X2180-1A;
 CC MEDLINE=91185402; PubMed=2010462;
 CC Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
 CC Yamasaki M.;
 CC "A cytoskeleton-related gene, usol, is required for intracellular
 CC RT protein transport in Saccharomyces cerevisiae.";
 CC J Cell Biol. 113:245-260(1991).
 CC [2]
 CC SEQUENCE OF 782-1790 FROM N.A.
 CC Hostetler M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
 CC Kendrick K.E.;
 CC Submitted (FEB-1993) to the EMBL/GenBank/DBJ databases.
 CC [3]
 CC SEQUENCE OF 1-8 FROM N.A.
 CC Bai Y., Symington L.S.;
 CC Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
 CC COMPLEX.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR

MEMBRANS. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
 ER AND THE GOLGI COMPLEX.
 -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, COMPOSED
 OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
 COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
 -1- SIMILARITY: BELONGS TO THE VDP/USOL/YB1047C FAMILY.

 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X54378; CA38253.1; -
 CC EMBL; L03188; AAB00143.1; -
 CC EMBL; U53668; AAB6655.1; -
 CC PTR; A38455; A38455.
 CC SGD; S0002216; USOL.
 CC DR InterPro; IPR002017; Spectrin.
 CC DR Transport; Protein transport; Golgi stack; Cytoskeleton; coiled coil.
 CC KM DOMAIN 1 724 GLOBULAR HEAD.
 CC FT DOMAIN 1 725 1790 COILED COIL (POTENTIAL).
 CC FT DOMAIN 465 487 CHARGED (HYPER-HYDROPHILIC).
 CC FT DOMAIN 991 1790 DISPENSABLE FOR THE PROTEIN FUNCTION.
 CC FT DOMAIN 1172 1786 ASP/GLU-RICH (ACIDIC).
 CC FT CONFLICT 847 847 E -> E (IN REF. 2).
 CC FT CONFLICT 924 924 G -> K (IN REF. 2).
 CC FT CONFLICT 1253 1253 V -> I (IN REF. 2).
 CC FT CONFLICT 1319 1319 I -> V (IN REF. 2).
 CC FT CONFLICT 1461 1461 N -> S (IN REF. 2).
 CC FT CONFLICT 1581 1581 G -> S (IN REF. 2).
 CC FT CONFLICT 1600 1600 I -> V (IN REF. 2).
 CC FT CONFLICT 1661 1661 R -> S (IN REF. 2).
 CC FT CONFLICT 1772 1772 D -> DEEDDEE (IN REF. 2).
 CC SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 6.8%; Score 127.5; DB 1; Length 1790;
 Best Local Similarity 19.7%; Pred. No. 0.89;
 Matches 72; Conservative 63; Mismatches 140; Indels 91; Gaps 14;

9 LSNDKRVNNINSEYKQYKLETOYATNDFLNLNBSKYKGLPG--KDIITF 65
 767 LTNHKEILD-----EKYQILNSSHSSLSKENSILETELKRVDRSIDEXTQLRDVLET 818
 66 NSADD-----KGIFRSEYCYCPDGVKCDGKITYHKSNDNREVNNDYRPMGVK 118
 819 KDKENOTALLEYKSTIKOE-----DSIKTLKG--LETTLSOKRAEDGIN 863
 119 PTNITVLYSGNEGDITOKLENFCSSTNYKDKNNQWECYK-----D 162
 864 KMGKDLFALSREM-----QAVENCKNLOKEDKKNVNHOKETKSLKDIAKITEIKAIN 919
 163 ENINRCKLEONT---ETNNDPKIISFHNPELWVYLLRDTI--KNNDKIKTCINNTTT 217
 920 ENLEEMKIQCNLSKKEHEHISKELVEYKSRQ-----SHDNLAKTEKLSKSIANN--- 970
 218 HCIDECNRNCLCFRWYKOREEWNISIKLFTKKNNIQOQSYNNINLFGY-----FF 271
 971 -----YKMDQAVENESLIKAVEESKNSSIQLSQVONKIDMSQSEKENFQ 1014
 272 KYMDKLDKDAKKEKLEMEIKRKNKE--FSNLENNRDYLENAIELLDHLEKETAATCKDN 330
 1015 IERSISIKKNIEQKLKTTISDLEQTKKEILISKSDSKDEYEQISLTKKL-ETATANDEN 1073
 331 TNEACE 336
 1074 VNKISE 1079

RESULT 8
 HMW2_MTCGE

ID HMW2_MYCCE STANDARD; PRT; 1805 AA.
AC P47460;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytadherence high molecular weight protein 2 (Cytadherence accessory protein 2).
GN HMW2 OR MG218.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhmann J.L.,
RA Nguyen D.T., Ufferback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bort K.F., Hu P.-C., Lucier T.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RA The minimal gene complement of Mycoplasma genitalium.";
RA Science 270:397-403(1995).
RN [2]
RP SEQUENCE OF 557-659 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bort K.F., Hutchison C.A. III,
RA "A survey of the Mycoplasma genitalium genome by using random
RT sequencing.";
RT J. Bacteriol. 175:7918-7930(1993).
CC -1- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
CC STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CC CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
CC PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHERESIN PROTEINS
CC IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
CC SIMILARITY).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: U39701; AAC71437.1;
CC EMBL: U02165; AAD12447.1;
CC TIGR: MG218;
CC
CC Cytadherence; Structural protein; Coiled coil; Complete proteome.
FT DOMAIN 28 838 COILED COIL (POTENTIAL).
FT DOMAIN 914 1591 COILED COIL (POTENTIAL).
FT DOMAIN 1632 1723 COILED COIL (POTENTIAL).
FT DOMAIN 1777 1804 COILED COIL (POTENTIAL).
SQ SEQUENCE 1805 AA; 216252 MW; 11D093AF173284FD CRC64;
Query Match 6.4%; Score 122; DB 1; Length 1805;
Best Local Similarity 20.5%; Pred. No. 2;
Matches 81; Conservative 57; Mismatches 121; Indels 136; Gaps 18;
OY 6 QSYLSNDNKKFVNINSEYKQFEKLEKQVATNTJFNLNEKYGKGLPGEEDIFT 65
DB 1234 QTYLANKRK-----AEY-----SQOOLQKYNLND-----LKENLERTD----- 1269
OY 66 NSADDKGFIFYSEYQV---CPDGVCKDGIKTKYKHSNDREVRNEDYKPPWGVKPTNI 122
DB 1270 -QLKRC--HHSIFARKLRFANDLFEKKOLKAQRIYDDKRLKLENER----- 1315
OY 123 TVLSGNGOGITOKLEFNCSTNYKDKNNK-----WECYIDENINCKLE-- 171
DB 1316 NMLHLSNF-----TERKRAVLDEQISYFEKQKQATDALILASHKEKKKEGLKILVELE 1371
OY 172 -QNTFNNNDNPKIISFNFFELWYLLDRFIKWNDKL-----TCINNTTTHC 219

DB 1372 TRTKTKLNDFAFKRSQREFE-----NORLLLEQTLQOTSSNFKTKA 1418
OY 220 IDECNRNCLCFQRMWYKQKEEWNISIKLPTKRNKIQOYSYNNINLFEQYFKVMDLKD 279
DB 1419 IOEINSS-----YKRGMEELNFOKKEPDKNK-----SRLYE-YFKKMDLEIR 1460
OY 280 DEAKKEKLEMNKRRKNES-----NLENNR-DYLENAIELLDHL----- 319
DB 1461 KESQVAVLEKFORANLLEQAANKLITEKNTIDREKELKAFKDKVDQIDSTNKORKE 1520
OY 320 -----KETATICKDNNTNFACEF 337
DB 1521 LNEILNENKLLQOSSLIERERAINSKDSLANKIET 1555
RESULT 9
ID PVDB_PLAKN STANDARD; PRT; 1153 AA.
AC P50493;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Duffy receptor, beta form precursor (Erythrocyte binding protein).
OS Plasmodium knowlesi.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5850;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92357776; PubMed=1496004;
RA Adams J.H., Sim B.K., Dolan S.A., Fang X., Kaslow D.C., Miller L.H.,
RA "A family of erythrocyte binding proteins of malaria parasites.";
RA Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
CC DETERMINANT.
CC
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC
CC -1- SIMILARITY: HIGH, TO P.VIVAX DUFFY RECEPTOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL: M90694; AAA29603.1;
CC
CC Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
KW MultiGene family.
FT SIGNAL 1 21
FT CHAIN 22 1153
FT DOMAIN 22 1105
FT TRANSMEM 1086 1106
FT DOMAIN 1107 1153
FT CARBOHYD 134 134
FT CARBOHYD 179 179
FT CARBOHYD 202 202
FT CARBOHYD 252 252
FT CARBOHYD 348 348
FT CARBOHYD 430 430
FT CARBOHYD 467 467
FT CARBOHYD 576 576
FT CARBOHYD 626 626
FT CARBOHYD 722 722
FT CARBOHYD 847 847
FT CARBOHYD 856 856
FT CARBOHYD 900 900
FT CARBOHYD 910 910
FT CARBOHYD 935 935
SQ SEQUENCE 1153 AA; 130471 MW; 6497BD1CE7BEC01 CRC64;
Query Match 6.3%; Score 121; DB 1; Length 1153;
Best Local Similarity 19.5%; Pred. No. 1.4;

```

RESULT 10
ARP_PLAFA ID ARP_PLAFA STANDARD; PRT; 537 AA.
AC P04931;
DT 13-AUG-1987 (Rel. 05, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Asparagine-rich protein (AG319) (ARP) (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
[1]
RN RP SEQUENCE FROM N.A.
RX MEDLINE=66176787; PubMed=2421257;
RA Stehl H.-D., Bianco A.E., Crewther P.E., Burkot T., Coppel R.L.,
RA Brown G.V., Anders R.F., Kemp D.J.;
RT "An asparagine-rich protein from blood stages of Plasmodium
RL falciparum shares determinants with sporozoites." ;
Nucleic Acids Res. 14:3089-3102(1986).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sdb.ch/announce/
or send an email to license@sdb.slb.ch).
-----
CC --
DR EMBL; M24328; AAA29491.1; -
DR PIR; A23770; A23770.
KW Malaria.
FT NON TER 1 1
FT TER 537 537
SO SEQUENCE 537 AA; 63000 MW; A6D76B15318CC239 CRC64;
-----
OY Query Match 6.3%; Score 120.5; DB 1; Length 537;
Best Local Similarity 19.3%; Pred. No. 0.65;
Matches 76; Conservative 73; Mismatches 156; Indels 89; Gaps 20;
DB 60 VEEHRRNNSSIDNNNSNNINNTYN----OOTRFS--SFEMENENENKNYHTGGM--NNN 109
OY 62 IFFTSADCKGFIFYSEKCYQVCPCDGVCND---GIKYFHKSNDNR---ERANN-EDVKP 113
DB 110 IHFKKKYIDNNNSSMKN-----TTNNKTIDISYNNKGITINDNNMMMDYLRNINNINEYA- 161

```

```

RESULT 11
Y091_CAEEL
ID Y091_CAEEL STANDARD; PRT; 782 AA.
AC Q10030;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical 90.1 kDa protein C27D6.1 in chromosome II.
GN C27D6.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditiida; Rhabditoidea;
OC Rhabditiidae; Peloderinae; Caenorhabditia.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol NZ;
RL Ding H.;
RA submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.lsb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
CC -----
DR EMBL; U23179; AAC6721.1; -.
DR WormPep; C27D6.1; CE01811.
KW Hypothetical protein; Colled coll.
FT DOMAIN 659 672 POLY-ALA.
FT DOMAIN 223 331 COLLED COLL (POTENTIAL).
FT DOMAIN 348 398 COLLED COLL (POTENTIAL).
FT DOMAIN 428 601 COLLED COLL (POTENTIAL).
FT DOMAIN 699 743 COLLED COLL (POTENTIAL).
SQ SEQUENCE 782 AA; 90059 MW; BB7CB842CD7273BB CRC64;

Query Match 6.2%; Score 119.5; DB 1; Length 782;
Best Local Similarity 21.1%; Pred. NO. 1.2; Mismatches 9; Gaps 9;
Matches 50; Conservative 40; Indels 78;

160 YK---DENINRCKLEQNTNEINNNDNPKRIISFHFNFELATWYLLLRDT----- 201
      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 181 YKRAGDAASMSRDLSPPSSVSFTTYLVAFHFKVPYAATYITREKRSERQTKVLESIER 240
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy 202 -IKMWDKLKTCTINNTTHTCIDEKNR-----CLCFDRW-VKOE----- 238
      ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 241 RLKANERKARKEIEBAEKWKMDRATRNSKRPLPELELAETVOAKEMQVKSQEIMQNKQ 300
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

Oy 239 --EENWSIKRLFTKKNNKIQQSYISI--NNLFESYEFKV-----MDKLIDDE 281
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Db 301 LVEELNEVOKKLEIENSOKTEFHOKVSTLNDIEEFQNPDEQDQSLQFMNDVLEQEI 360
 QY 282 ANKMELENIRK-----KNFSNLNNRYLNALELDLH-----KEATATCK 327
 Db 361 VAYQECISLKQENELKELKLOQLSSLTVMNHVSTLMDHLEINKQESRETOGICK 417

RESULT 12
 Y228_BORBU STANDARD; PRT; 971 AA.
 AC 051246;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein BB0228.
 OS BB0228.
 GN Borrelia burgdorferi (Lyme disease spirochete).
 OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
 NCBI_TaxID=139;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=ATCC 35210 / B31;
 MEDLINE=98065943; PubMed=9403685;
 RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
 RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
 RA Peterson J., Tomb J.F., Fleischmann R.D., Richardson D.,
 RA van Vugt R., Palmer A.R., Adams M.D., Quackenbush J., Salzberg S., Hanson M.,
 RA Ustergaard T., Wathey L., McDonald L., Artach P., Bowman C.,
 RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
 RA Smith H.O., Venter J.C.;
 RT "Genomic sequence of a Lyme disease spirochete, Borrelia
 burgdorferi."
 RL Nature 380:580-586(1997).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AEO01133; AAC66621.1;
 DR TIGR; BB0228;
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 971 AA; 112959 MW; 088A68BD7B8C591A CRC64;

Query Match 6.2%; Score 119; DB 1; Length 971;
 Best Local Similarity 23.1%; Pred. No. 1.6;
 Matches 82; Conservative 42; Mismatches 121; Indels 110; Gaps 19;

QY 4 ELQSYLNDKRFVNNINSEYKOFYEKLEKTOYATNDTFLNLNEG-KYCGGLDGE-K 60
 Db 156 EMGGSYSNNKSLINEYSS-----SLEEGVAYYDGGGDIPTNII 194
 QY 61 DITFTNSADDKGIFRSEYQVCPDGVKCDGKIKYKSDNRERNNDYKPPMGVKT 120
 Db 195 DLYVESFLD-----FYKKY--TLENCKIFLPGNTQTEKMLNTEKTIIRPYKE----- 242
 QY 121 NITVYLSNGEGDITQKLENFQNSSTNYKDKNNQKE-----CYR--DENNR----- 167
 Db 243 -----KSNVNINIEVKKRWEKGRKLTJKIPKENDNSLGVTIN 280
 QY 168 --CKLEQNTENNNDPKIISPHNFELWYTLRDITKWN-DKLTCTCINNTHHCIDECN 224
 Db 281 WIC-----TEINNIDST-----GEILSEILLDSCSTTINILKSGIEDIAH-ISGIN 329
 QY 225 RN-----CLCPDRVKKQKEEEMSIKILFTKKNKIOOSYYSN--INMLFEGYFKVWDK 276
 Db 330 TDLKESIFSGIQNVVENEKEFKNL--VFSEKLVKVNKIKPKELIKGLIGYEYALKEE 387
 QY 277 -----LDKDEAKMELENIRK-----NEFSN-LENNRDYLENAIELL 316

Db 388 KGQNFPLAMIKSRKGNLGHPIKTLQTSYTYINEITNKEKIYYFENLEIYTL 442

RESULT 13
 PYDA_PLAKN STANDARD; PRT; 1073 AA.
 AC P22345;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Duffy receptor alpha form precursor (Erythrocyte binding protein).
 DE Plasmodium knowlesi.
 OS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=5850;
 [1]
 SEQUENCE FROM N.A.
 MEDLINE=92357776; PubMed=1496004;
 RA Adams J.H., Sim B.K., Dolan S.A., Pang X., Kaslow D.C., Miller L.H.,
 RA Proc. Natl. Acad. Sci. U.S.A. 89:7085-7089(1992).
 [2]
 SEQUENCE OF 296-1073 FROM N.A.
 MEDLINE=91004213; PubMed=2170017;
 RA Adams J.H., Hudson D.E., Torii M., Ward G.E., Wellens T.E.,
 RA Aikawa M., Miller L.H.;
 RT "The Duffy receptor family of Plasmodium knowlesi is located within
 RT the micromeres of invasive malaria merozoites."
 RL Cell 63:141-153(1990).
 CC -1- FUNCTION: BINDS TO THE HUMAN ERYTHROCYTES DUFFY BLOOD GROUP
 CC DETERMINANT.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: HIGH TO P.VIVAX DUFFY RECEPTOR.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; M90466; AAA29602.1;
 DR EMBL; M68517; AAA29590.1;
 DR EMBL; M68518; AAA29591.1;
 DR PIR; A35970; A35970.
 KW Malaria; Receptor; Glycoprotein; Signal; Transmembrane;
 KW Multigene family.
 FT SIGNAL 1 21
 FT CHAIN 1 21
 FT DOMAIN 22 1073
 FT TRANSMEM 1008 1029
 FT DOMAIN 1030 1073
 FT CARBOHYD 134 134
 FT CARBOHYD 179 179
 FT CARBOHYD 202 179
 FT CARBOHYD 252 202
 FT CARBOHYD 348 348
 FT CARBOHYD 679 679
 FT CARBOHYD 746 746
 FT CARBOHYD 779 779
 FT CARBOHYD 788 788
 SQ SEQUENCE 1073 AA; 120683 MW; 3965FC9F46B71808 CRC64;

Query Match 6.2%; Score 119; DB 1; Length 1073;
 Best Local Similarity 20.0%; Pred. No. 1.8;
 Matches 81; Conservative 64; Mismatches 149; Indels 110; Gaps 18;

QY 7 SYLSDNKKFVNNINSEYKOFYEKLEKTOYATNDTFL--NLNE--GKYCKG-----GLPG 58
 Db 175 NPLENSK-----GQHPYRNKRERKSNVGINOTFLQNNVNDKCDKRRGRBDMDCA 229
 QY 59 EKDI-----FTNSADDKGIFRSEYQVCPDGVKCDGKIKYTHK 98

```

Db 230 EXDICSIVRRYOLCMGLTNLVNTRTHSHNDTFLKLNKRLKMLDAVAEGDILL- 285
OY 99 SDDNREYVNDKYP-----MGVKPTNITVYSGNEGDIOQKLEFNCNS--STNYKD 150
Db 286 -----KKNYQYNKECKAIRKGLDGFIDITIMKTNEGIGYSOVANNRQVFGDIKA 339
OY 151 KNNOK-----WECY-----YKDENINRCKLEONTEINNDPKIISFHEFELW 193
Db 340 KORKQWNNSSKEHIMWAMFIRSRLKEKFWIKCKDVLKY---EPQI---YRMIREW 393
OY 194 ----VYLLADDTIKNDKL--KTCINNTTTCIDECNRNCLCDEPRWYKQKEEENSTK 247
Db 394 GRDYSKLPKEQGLNKCKASKLYNNMAICMPLCHDACKSYDOWITRRKKQMDVLSTK 453
OY 248 FTKKKNIOQSYNINNLFGYFFKVMKDILDKDEAKKKELEMINIKRKNNEPSLNENRDY 307
Db 454 FSSYKTKQKGTENTIAATY-----DILKQELNGFEK-----TFENIENKRDN 496
308 LENAIELLDHLKETATICKDNNTNEACET-----SHNATNP 345
497 LYN-----HLCPCVVEARKNTQENVKNGSGVESKASSNP 533

```

```

RESULT 14
IFH1_YEAST STANDARD; PRT; 1085 AA.
AC P39520;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE IFH1 protein (RRP3 protein).
GN IFH1 OR RRP3 OR YLR223C OR L8083.9.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycos.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=FL100;
RX MEDLINE=95304839; PubMed=7785326;
RA Cherel I., Thuriaux P.;
RT "The IFH1 gene product interacts with a fork head protein in
RL yeast 11:261-270(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=5288c / AB972;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favell A., Fulton L., Gatlung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Mardis E., Meneses S., Miller N., Nhan M., Pauley A., Peluso D.,
RA Rifken L., Riles L., Taich A., Trevasakis E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CONTROLS THE PRE-RNA PROCESSING MACHINERY IN
CC CONJUNCTION WITH IFH1. COULD CONVERT IFH1 FROM A REPRESSOR
CC TO AN ACTIVATOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: Z29488; CA82624.1;
CC DR EMBL: U19027; AAB67412.1;
CC DR PIR: S47477; S47477.
CC SGP: S0004213; IFH1.

```

```

KW Nuclear protein; Transcription regulation.
FT DOMAIN 122 163 ASP/GLU-RICH (HIGHLY ACIDIC).
SQ SEQUENCE 1085 AA; 122491 MW; BEIC/DEF06213FE0 CRC64;
Query Match 6.2%; Score 118.5; DB 1; Length 1085;
Best Local Similarity 20.3%; Pred. No. 2;
Matches 86; Conservative 66; Mismatches 153; Indels 119; Gaps 15;
OY 10 SDDNRFY-----NNINSEYKQFYERKLEKTOYATNDTFLNLNBEKCYKGLPGEKDITFT 65
Db 199 SNKKSQVSLSPKRENEDEQKEKEKEKEQKQESKKEVNG--SGTTTQOALSPK 254
OY 66 NSADDKGIFTRSEYCYQPCGKVC-----DGIYTHKSD 100
Db 255 FKEDDDISFGNGEGYNEIDGEEVLDLKNKNNNGNEEDKLSKVMGLNDELFPNISE 314
OY 101 NDRER-----VNNEDYKRPWGKPTNITVLVSG-----NEGDIOQKLEN- 140
Db 315 SDESEYIDODAYFDVYNNEDSHGEIG-----TDLEFGEDDLPILEBEQNTVSEIQND 368
OY 141 ----FCNS-----STNYKDKNNQKWEQYKDNINRCKLEQNTIIND-----NPKII 184
Db 369 DELSFGDSIHESGDPYEDAKNFKLQNEVNGENGYDEEDDEDEIDMSDFMPYEDKFA 428
OY 185 SFHNFELWYTYLLRDITKNDKLTCTINNTTTCIDECNRNCLCFDRWYKQKEEENSTI 244
Db 429 NLYYTGD-----GSEPKLSLSTSLPLMLNDELKSLKKEAKKREDEERROR 475
OY 245 KKLFTKKNIOQSYNINNLFGYFFKVMKDILDKDEAKKKELEMINIKRKNNEF 298
Db 476 RKLTKKQKQESTRTTSVND--DEYINVFQSDENSGHKSCKGRHKSGLHLEHKKG 533
OY 299 SNLENNRDYLENAI-----ELLDLHLEKETATICKDNNTNEACETSHN 340
Db 534 SNLKNSDDLEPRSHSYVNLNGSKYDSSDDEYDNLILD---VAHMPDSECS-ETSHD 588
OY 341 ATTN 344
Db 589 ADTD 592
RESULT 15
IF2P_HUMAN STANDARD; PRT; 1220 AA.
AC O60841; O95805; Q9UFB1; Q9UAM7;
DT 15-DEC-1998 (Rel. 37, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Translation Initiation Factor IF-2.
GN IF2 OR KIA0741.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RL for large proteins in vitro."
RN [2]
RP SEQUENCE FROM N.A.
RA TISSUE=Cervical carcinoma;
RX MEDLINE=99362399; PubMed=10432305;
RA Wilson S.A., Sleiro-Vazquez C., Edwards N.J., Iourin O., Byles E.D.,
RA Kotsopoulos E., Adams C.S., Kingsman S.M., Kingsman A.J.,
RA Martin-Rendon E.;
RT "Cloning and characterization of hIF2, a human homologue of bacterial
RT translation initiation factor 2 and its interaction with HIV-1

```

```

RT matrix." ;
RL Biochem. J. 342:97-103(1999).
[3]
SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Testis;
RX MEDLINE=99218282; PubMed=10200264;
RA Lee J.H., Choi S.K., Roll-Meck A., Burley S.K., Dever T.E.;
RT "Universal conservation in translation initiation revealed by human
RF and archaeal homologs of bacterial translation initiation factor
R1 IF2." ;
RL Proc. Natl. Acad. Sci. U.S.A. 96:4342-4347(1999).
[4]
SEQUENCE OF 89-1220 FROM N.A.
RP TISSUE=Testis;
RA Kohrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (DEC 1999) to the EMBL/Genbank/DBJ databases.
[5]
SEQUENCE OF 833-1220 FROM N.A.
RX MEDLINE=21064499; PubMed=11124703;
RA Stanchi F., Bertocco E., Toppo S., Diognardi R., Simlonati B.,
RA Cannata N., Zimballo R., Lanfranchi G., Valle G.;
RL "Characterization of 16 novel human genes showing high similarity to
yeast sequences." ;
Yeast 18:69-80(2001).
-1- FUNCTION: FUNCTION IN GENERAL TRANSLATION INITIATION BY PROMOTING
THE BINDING OF THE FORMYL METHIONINE--TRNA TO RIBOSOMES. SEEMS TO
FUNCTION ALONG WITH EIF-2 (BY SIMILARITY).
-1- SIMILARITY: BELONGS TO THE IF-2 FAMILY.
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL; AB018284; BAA34461.1; -
DR EMBL; AJ006776; CAB44357.1; -
DR EMBL; AF078035; AAD16006.1; -
DR EMBL; AL133563; CAB63717.1; -
DR EMBL; AJ006412; CAA07018.1; -
MIM; 606086; -.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR000178; IF2.
DR InterPro; IPR005225; Small_GTP.
Pfam; PF00009; GTP_EFTU; 1.
Pfam; PF03144; GTP_EFTU_D2; 2.
PRINTS; PR00315; ELONGFACT.
TIGRfams; TIGR00231; small_gtp; 1.
DR PROSITE; PS01176; IF2; FALSE_NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding.
FT NP_BIND 638 645 GTP (BY SIMILARITY).
FT DOMAIN 39 50 POLY-LYS.
FT DOMAIN 94 99 POLY-LYS.
FT DOMAIN 138 142 POLY-ASP.
FT DOMAIN 313 322 POLY-LYS.
FT DOMAIN 353 356 POLY-LYS.
FT DOMAIN 361 364 POLY-GLU.
FT DOMAIN 491 496 POLY-GLU.
FT DOMAIN 529 567 ASP/GLU-RICH (ACIDIC).
FT MUTAGEN 640 640 V->G; LOSS OF ACTIVITY IN VIVO. RETAINS
FULL ACTIVITY IN VITRO.
FT MUTAGEN 706 706 H->E; LOSS OF ACTIVITY; BOTH IN VIVO AND
IN VITRO.
FT MUTAGEN 706 706 H->Q; LOSS OF ACTIVITY IN VIVO. PARTIAL
ACTIVITY IN VITRO.
FT MUTAGEN 706 706 D->N; LOSS OF ACTIVITY; BOTH IN VIVO AND
IN VITRO.
FT MUTAGEN 759 759 E -> G (IN REF. 1).
FT CONFLICT 64 64 T -> I (IN REF. 2).
FT CONFLICT 92 92 I -> M (IN REF. 3).
FT CONFLICT 180 180

```

[illegible]

Search completed: April 28, 2003, 10:30:19
Job time : 13.3149 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 12.9282 Seconds

(without alignments)
5498.552 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1918
Sequence: 1 YEKETQSYLSNDNKFVNININ.....CRDNTNACETSHNATNP 345Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

al number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1918	100.0	3542	5	Q9U5M2
2	1864	97.2	510	5	Q8T6L0
3	1766	92.1	492	5	Q8T6K9
4	1765	92.0	494	5	Q8T6K7
5	1718	89.6	1615	5	Q8T325
6	1717	89.5	3287	5	Q8T325
7	1631	85.0	461	5	Q8T6K4
8	1625	84.7	461	5	Q8T6K3
9	1610	83.9	492	5	Q8T6K6
10	1609	83.9	494	5	Q8T6K0
11	1603	83.6	496	5	Q8T6K8
12	1577	82.2	494	5	Q8T6K5
13	1465	76.4	460	5	Q8T6K1
14	1260	65.7	427	5	Q8T6K2
15	771	40.2	3006	5	Q26032
16	513	26.7	1685	5	Q9U4A2

17	501.5	26.1	2647	5	P90580	P90580	plasmodium
18	499.5	26.0	2212	5	Q94657	Q94657	plasmodium
19	457	23.8	2664	5	Q26033	Q26033	plasmodium
20	451	23.5	2169	5	Q97312	Q97312	plasmodium
21	446	23.3	2209	5	Q9U0G6	Q9U0G6	plasmodium
22	446	23.3	2228	5	Q60991	Q60991	plasmodium
23	444	23.1	2270	5	Q9XZB8	Q9XZB8	plasmodium
24	442.5	23.1	2527	5	Q95W83	Q95W83	plasmodium
25	437	22.8	3026	5	Q26030	Q26030	plasmodium
26	432.5	22.5	2197	5	Q96296	Q96296	plasmodium
27	429	22.4	2277	5	Q9U0G5	Q9U0G5	plasmodium
28	407	22.2	480	5	Q9NAV6	Q9NAV6	plasmodium
29	403.5	21.0	2706	5	Q15870	Q15870	plasmodium
30	399.5	20.8	2658	5	Q8T5G0	Q8T5G0	plasmodium
31	394.5	20.6	3078	5	Q26031	Q26031	plasmodium
32	380	19.8	2182	5	Q26034	Q26034	plasmodium
33	363	18.9	1711	5	Q96108	Q96108	plasmodium
34	362	18.9	2209	5	Q97324	Q97324	plasmodium
35	361.5	18.8	2135	5	Q61077	Q61077	plasmodium
36	357.5	18.6	1729	5	Q25734	Q25734	plasmodium
37	357.5	18.6	2924	5	Q25733	Q25733	plasmodium
38	356	18.6	2163	5	Q9NFB6	Q9NFB6	plasmodium
39	331	17.3	2042	5	Q25766	Q25766	plasmodium
40	231	12.0	921	5	Q25989	Q25989	plasmodium
41	223.5	11.7	174	5	Q15664	Q15664	plasmodium
42	223.5	11.7	174	5	Q18671	Q18671	plasmodium
43	219	11.4	130	5	Q15662	Q15662	plasmodium
44	216.5	11.3	174	5	Q15652	Q15652	plasmodium
45	216.5	11.3	174	5	Q15665	Q15665	plasmodium

ALIGNMENTS

RESULT 1

Q9U5M2 PRELIMINARY; PRT; 3542 AA.
AC Q9U5M2;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)
DE PCR3 CSA ligand (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NC NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PCR3;
RX MEDLINE=20006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheidig C., Baruch B., Oishi S., Fujii N.,
RA Fusai T., Parzy D., Miller L.H., Gysin J., Scherf A.:
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
RT sulfate A: A receptor for human placental infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).
DR EMBL: AJ13811; CAB59840.1;
DR InterPro: IPR004258; Neurotoxin.
DR InterPro: IPR004258; PEMP.
DR Pfam: PF03011; PEMP.1.
DR PRINTS: PR00284; TOXIN.
FT NON_TER 3542 3542
SQ SEQUENCE 3542 AA; 413089 MW; 970D85EE88DA2EC2 CRC64;

Query Match 100.0%; Score 1918; DB 5; Length 3542;

Best Local Similarity 100.0%; Pred. No. 7.5e+108; Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEKETQSYLSNDNKFVNINSEYKQFEKLETOYATNDFNLNMGKCYCKGLPEK 60
DB 403 YEKETQSYLSNDNKFVNINSEYKQFEKLETOYATNDFNLNMGKCYCKGLPEK 462
QY 61 DITFNSADDKGIRYRSYCYCPCGCKGICGKITKTHSDNDREVRNEDKRPNGVAPT 120
|||||

DB 463 DITFTNSADDKGIFYYSEYQVCPDGVKDGKIKYTHKSDNDRERANNEDYKPPMGVCKPT 522
 QY 121 NITVLYSGNEQGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 180
 DB 523 NITVLYSGNEQGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 562
 QY 181 PRTISPHNFEFLWVYLLRDTIKMNDKLTCTCINNNTTHCIDEENRNCLECFDRWVYKKEE 240
 DB 583 PKTISPHNFEFLWVYLLRDTIKMNDKLTCTCINNNTTHCIDEENRNCLECFDRWVYKKEE 642
 QY 241 WNSIKKLFTRKKNNIOOSYSYNNINLEFEGYFFVYMDKLDKDEAKWELMENIKRKKNEFSN 300
 DB 643 WNSIKKLFTRKKNNIOOSYSYNNINLEFEGYFFVYMDKLDKDEAKWELMENIKRKKNEFSN 702
 QY 301 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHNATNP 345
 DB 703 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHNATNP 747

RESULT 2
 6LO
 08T6L0 PRELIMINARY; PRT; 510 AA.

AC 08T6L0;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy."
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL: AF45886; AAL84269.1; -.
 FT NON_TER 1
 FT NON_TER 510
 FT SEQUENCE 510 AA; 60713 MW; E3CF55DC04601099 CRC64;

Query Match 97.2%; Score 1864; DB 5; Length 510;
 Best Local Similarity 98.5%; Pred. No. 1.9e-105;
 Matches 335; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YKEKIOSYLSNDKKEFVNNINSEYKOFYKELKETOYATNDTFLINLENGKYGKGLPGER 60
 DB 171 YKEKIOSYLSNDKKEFVNNINSEYKOFYKELKETOYATNDTFLINLENGKYGKGLPGER 230
 DB 61 DITFTNSADDKGIFYYSEYQVCPDGVKDGKIKYTHKSDNDRERANNEDYKPPMGVCKPT 120
 QY 121 NITVLYSGNEQGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 180
 DB 291 NITVLYSGNEQGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 350
 QY 181 PRTISPHNFEFLWVYLLRDTIKMNDKLTCTCINNNTTHCIDEENRNCLECFDRWVYKKEE 240
 DB 351 PKTISPHNFEFLWVYLLRDTIKMNDKLTCTCINNNTTHCIDEENRNCLECFDRWVYKKEE 410
 QY 241 WNSIKKLFTRKKNNIOOSYSYNNINLEFEGYFFVYMDKLDKDEAKWELMENIKRKKNEFSN 300
 DB 411 WNSIKKLFTRKKNNIOOSYSYNNINLEFEGYFFVYMDKLDKDEAKWELMENIKRKKNEFSN 470
 QY 301 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHN 340
 DB 471 LENNRDYLENAIEILLDHLKETATICKDNTNACETSHN 510

RESULT 3
 08T6K9

ID 08T6K9 PRELIMINARY; PRT; 492 AA.
 AC 08T6K9;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy."
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL: AF45886; AAL84270.1; -.
 FT NON_TER 1
 FT NON_TER 492
 FT SEQUENCE 492 AA; 58590 MW; 277204FDF55014C6 CRC64;

Query Match 92.1%; Score 1766; DB 5; Length 492;
 Best Local Similarity 97.5%; Pred. No. 1.6e-99;
 Matches 316; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 1 YKEKIOSYLSNDKKEFVNNINSEYKOFYKELKETOYATNDTFLINLENGKYGKGLPGER 60
 DB 169 YKEKIOSYLSNDKKEFVNNINSEYKOFYKELKETOYATNDTFLINLENGKYGKGLPGER 228
 QY 61 DITFTNSADDKGIFYYSEYQVCPDGVKDGKIKYTHKSDNDRERANNEDYKPPMGVCKPT 120
 DB 229 DITFTNSADDKGIFYYSEYQVCPDGVKDGKIKYTHKSDNDRERANNEDYKPPMGVCKPT 288
 QY 121 NITVLYSGNEQGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 180
 DB 289 NITVLYSGNEQGDITOKLENFCNSSTNYKDKNNQKWECCYKDEININCKLEQTEINNDN 348
 QY 181 PRTISPHNFEFLWVYLLRDTIKMNDKLTCTCINNNTTHCIDEENRNCLECFDRWVYKKEE 240
 DB 349 PKTISPHNFEFLWVYLLRDTIKMNDKLTCTCINNNTTHCIDEENRNCLECFDRWVYKKEE 408
 QY 241 WNSIKKLFTRKKNNIOOSYSYNNINLEFEGYFFVYMDKLDKDEAKWELMENIKRKKNEFSN 300
 DB 409 WNSIKKLFTRKKNNIOOSYSYNNINLEFEGYFFVYMDKLDKDEAKWELMENIKRKKNEFSN 468
 QY 301 LENNRDYLENAIEILLDHLKETAT 324
 DB 469 LENNRDYLENAIEILLDHLKETAT 492

RESULT 4
 08T6K7
 ID 08T6K7 PRELIMINARY; PRT; 494 AA.
 AC 08T6K7;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 RT implicated in malaria in pregnancy."
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL: AF45886; AAL84272.1; -.
 FT NON_TER 1
 FT NON_TER 494
 FT SEQUENCE 494 AA; 58773 MW; C6D6731157A1C13A CRC64;

Query Match 92.0%; Score 1765; DB 5; Length 494;
 Best Local Similarity 96.6%; Pred. No. 1.9e-99;
 Matches 315; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 YEKEIOSYATSNDRKFNININSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 60
 DB 169 YEKEIOSYATSNDRKFNININSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 228
 QY 61 DIFTNSADBDKGFYRSEYCOVCPDCGVKCDGKITYTHKSDNDRERNNEDYKPPMGVKPT 120
 DB 229 DIFTNSADBDKGFYRSEYCOVCPDCGVKCDGKITYTHKSDNDRERNNEDYKPPMGVKPT 288
 QY 121 NITVLVSGNEQDITQKLENFCSSTNYKDKNNQKWECCYKDKENINRCKLEQNTTEINNDN 180
 DB 289 NITVLVSGNEQDITQKLENFCSSTNYKDKNNQKWECCYKDKENINRCKLEQNTTEINNDN 348
 QY 181 PKTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRWVKOKEE 240
 DB 349 PKTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRWVKOKEE 408
 QY 241 WNSIKKLFTRKKNIQOSYSNINNLFEGYFFKVMYDKLDKDEAKWELMENIKRKKNEFSN 300
 DB 409 WNSIKKLFTRKKNNVPQYPTTININLFEGYFFKVMYDKLDKDEAKWELMENIKRKKNEFSN 468
 QY 301 LENNRDYLENAIELLDHLKETATTC 326
 DB 469 LENNRDYLENAIELLDHLKETATTC 494

RESULT 5

Q8T325 PRELIMINARY; PRT; 1615 AA.

AC Q8T325; 08T325;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE PFEMP1 (Fragment).
 GN TM284VAR3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TM284;
 RX MEDLINE-21927235; PubMed-11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 "Identification of a conserved Plasmodium falciparum var gene
 implicated in malaria in pregnancy."
 J. Infect. Dis. 185:1207-1211(2002).
 EMBL; AJ420412; CAD20868.1; -.
 FT NON_TER 1615 1615
 SQ SEQUENCE 1615 AA; 188639 MW; FCA896C00B86DEA6 CRC64;

Query Match 89.6%; Score 1718; DB 5; Length 1615;
 Best Local Similarity 89.3%; Pred. No. 4.3e-96;
 Matches 308; Conservative 14; Mismatches 23; Indels 0; Gaps 0;

QY 1 YEKEIOSYATSNDRKFNININSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 60
 DB 393 YEKEIOSYATSNDRKFNININSEYKOFHDLRKNKKNLDTFLNLNEGKCGGLPGEK 452
 QY 61 DIFTNSADBDKGFYRSEYCOVCPDCGVKCDGKITYTHKSDNDRERNNEDYKPPMGVKPT 120
 DB 453 DIFTNSADBDKGFYRSEYCOVCPDCGVKCDGKITYTHKSDNDRERNNEDYKPPMGVKPT 512
 QY 121 NITVLVSGNEQDITQKLENFCSSTNYKDKNNQKWECCYKDKENINRCKLEQNTTEINNDN 180
 DB 513 NITVLVSGNEQDITQKLENFCSSTNYKDKNNQKWECCYKDKENINRCKLEQNTTEINNDN 572
 QY 181 PKTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRWVKOKEE 240
 DB 573 PKTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRWVKOKEE 632

QY 241 WNSIKKLFTRKKNIQOSYSNINNLFEGYFFKVMYDKLDKDEAKWELMENIKRKKNEFSN 300
 DB 633 WNSIKKLFTRKKNNVPQYPTTININLFEGYFFKVMYDKLDKDEAKWELMENIKRKKNEFSN 692

QY 301 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHATNP 345
 DB 693 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHATNP 737

RESULT 6

Q8T326 PRELIMINARY; PRT; 3287 AA.

AC Q8T326; 08T326;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
 DE PFEMP1 (Fragment).
 GN TM180VAR2.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TM180;
 RX MEDLINE-21927235; PubMed-11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 "Identification of a conserved Plasmodium falciparum var gene
 implicated in malaria in pregnancy."
 J. Infect. Dis. 185:1207-1211(2002).
 EMBL; AJ420411; CAD20867.1; -.
 FT NON_TER 3287 3287
 SQ SEQUENCE 3287 AA; 383550 MW; 58F8F866FC244536 CRC64;

Query Match 89.5%; Score 1717; DB 5; Length 3287;
 Best Local Similarity 89.0%; Pred. No. 1e-95;
 Matches 307; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

QY 1 YEKEIOSYATSNDRKFNININSEYKOFYEKLETOYATNDTFLNLNEGKCGGLPGEK 60
 DB 393 YEKEIOSYATSNDRKFNININSEYKOFHDLRKNKKNLDTFLNLNEGKCGGLPGEK 452
 QY 61 DIFTNSADBDKGFYRSEYCOVCPDCGVKCDGKITYTHKSDNDRERNNEDYKPPMGVKPT 120
 DB 453 DIFTNSADBDKGFYRSEYCOVCPDCGVKCDGKITYTHKSDNDRERNNEDYKPPMGVKPT 512
 QY 121 NITVLVSGNEQDITQKLENFCSSTNYKDKNNQKWECCYKDKENINRCKLEQNTTEINNDN 180
 DB 513 NITVLVSGNEQDITQKLENFCSSTNYKDKNNQKWECCYKDKENINRCKLEQNTTEINNDN 572
 QY 181 PKTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRWVKOKEE 240
 DB 573 PKTISFHNFEFLWVYLLRDTIKMNDKLTCTINNTHHCIDECNRNCLCFDRWVKOKEE 632
 QY 241 WNSIKKLFTRKKNIQOSYSNINNLFEGYFFKVMYDKLDKDEAKWELMENIKRKKNEFSN 300
 DB 633 WNSIKKLFTRKKNNVPQYPTTININLFEGYFFKVMYDKLDKDEAKWELMENIKRKKNEFSN 692
 QY 301 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHATNP 345
 DB 693 LENNRDYLENAIELLDHLKETATTCNDNTNACETSHATNP 737

RESULT 7

Q8T6K4 PRELIMINARY; PRT; 465 AA.

AC Q8T6K4; 08T6K4;
 DT 01-JUN-2002 (Tremblrel. 21, Created)
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5683;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MS7;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RT implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
EMBL; AF458892; AAL84275.1; -.
FT NON_TER 1 465
FT NON_TER 1 465
SQ SEQUENCE 465 AA; 55496 MW; 60BE35F62468A37B CRC64;

Query Match	85.0%;	Score 1631;	DB 5;	Length 465;
Best Local Similarity	98.0%;	Pred. No. 2.3e-91;		
Matches 290;	Conservative 5;	Mismatches 1;	Indels 0;	Gaps 0;

[illegible]

RESULT 8			
ID	Q8T6K3	PRELIMINARY;	PRT; 461 AA.
AC	Q8T6K3;		
DT	01-JUN-2002 (TREMBLrel. 21, Created)		
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Erythrocyte membrane protein 1 (fragment). VAR. Plasmodium falciparum. Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium. NCBI_TaxID=5833;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN-M58;		
RC	RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;		
RT	"Identification of a conserved Plasmodium falciparum var gene implicated in malaria in pregnancy.";		
RL	J. Infect. Dis. 0:0-0(2002).		
DR	EMBL; AF458893; AAL84276.1; -.		
FT	NON_TER	1	1
FT	NON_TER	461	461
SO	SEQUENCE	461 AA;	54946 MW; F4C8863D50DA50EF CRC64;

Query Match	84.7%;	Score 1625;	DB 5;	Length 461;
Best Local Similarity	98.0%;	Pred. No. 5.2e-91;		
Matches 289;	Conservative 5;	Mismatches 1;	Indels 0;	Gaps 0;

QY 1 YEKELQSLSDNKNFVNNINSEYKKQFEKLEKETOYAINDPFLNLNGKCYCKGGLPGEK 60
 |||||
 Db 167 YEKELIESVSDAKFVNNINSEYKKQFEKLEKETOYAINDPFLNLNGKCYCKGGLPGEK 226
 |||||
 QY 61 DIFETNSADDKGITRYSRCQVCPDGCYKCGGIKYTHNSDDREVNNEDEYKPPMGVAPT 120
 |||||

Db	227	DIIFNNSADMDGIIYRSTRYQVYCDGCKDGIKYYTHKSDNDRVRVNNEDIKPPWGYKPT	266
Qy	121	NITVLYSGNEGDDITQKLENFCSNSTYKOKNOKMWCYKADENINCKLEONTEIINDN	180
Db	287	NITVLYSGNEGDDITQKLENFCSNSTYKOKNOKMWCYKADENINCKLEONTEIINDN	346
Qy	181	KPIISFNHFFELMWTYILLRDTIKKNDKLCITINNTHHCIDECKNRNLCEDRYVYKQKEE	240
Db	347	KPIISFNHFFELMWTYILLRDTIKKNDKLCITINNTHHCIDECKNRNLCEDRYVYKQKEE	406
Qy	241	WNSIKKLETTKKRNKOQSYYSINNLFEGYFPKVDKLDXDEAKMKELMEYIKRRK	295
Db	407	WNSIKKLETTKKRNKOQSYYSINNLFEGYFPKVDKLDXDEAKMKELMEYIKRRK	461

RESULT 9	
Q8T6K6	
ID Q8T6K6	PRELIMINARY;
AC Q8T6K6	PRT; 492 AA.

DT	01-JUN-2002	(TREMBLrel. 21, Created)
DT	01-JUN-2002	(TREMBLrel. 21, Last sequence update)
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)
DE	Erythrocyte membrane protein 1 (fragment).	
GN	Var.	
OS	Plasmodium falciparum.	
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
OX	NCBI_TaxID=56833;	
RN	[1]	
RP	SEQUENCE FROM N.A..	
RC	STRAIN=M48;	
RA	Rove J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;	
RT	"Identification of a conserved Plasmodium falciparum var gene	
RT	implicated in malaria in pregnancy.";	
RL	J. Infect. Dis. 0:0-0(2002).	
DR	EMBL: AF458890; AAL84273.1; -.	
FT	NON_TER	1
FT	NON_TER	1
FT	NON_TER	1
SO	SEQUENCE	492 AA; 56746 MW; 49D385FF0914B34A CRC64;

Query Match	83.98;	Score 1610;	DB 5;	Length 492;
Best Local Similarity	89.28;	Pred. No. 4.5e-90;		
Matches 288;	Conservative 17;	Mismatches 18;	Indels 0;	Gaps 0;

[illegible]

RESULT	10
08T6K0	
ID	Q8T6K0
AC	Q8T6K0;
DT	01-JUN-2002 (Tremblrel, 21, Created)
	PRELIMINARY;
	PRT;
	494 AA.

DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
GN Erythrocyte membrane protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN NCBI_TaxID=5833;
RP SEQUENCE FROM N.A.
RC STRAIN=VAN34;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implied in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL: AF458896; AAL84279.1; -.
FT NON_TER 1 1
FT NON_TER 494 494
SQ SEQUENCE 494 AA; 59240 MW; 739BE121652A367B CRC64;

Query Match 83.9%; Score 1609; DB 5; Length 494;
Best Local Similarity 88.9%; Pred. No. 5, 2e-90;
Matches 288; Conservative 14; Mismatches 22; Indels 0; Gaps 0;

OY 1 YEKEIOSYLSNDKFNANNSEYKQFEKLEQYATNDFLNLNEGKCKGLPGEK 60
DB 171 YEKEIOSYLSNDKFNANNSEYKQFEKLEQYATNDFLNLNEGKCKGLPGEK 230
OY 61 DIFTNSADKGFYRSEYQVCPDGVKCDGKITYHKSNDREVRNEDYKPPWYKPT 120
DB 231 DIFTNSADKGFYRSEYQVCPDGVKCDGKITYHKSNDREVRNEDYKPPWYKPT 290
OY 121 NITVLYSGNQGDIQKLEFNCNSSTYKDKNNQKWCYKDEINCKLEQNTIINN 180
DB 291 NITVLYSGNQGDIQKLEFNCNSSTYKDKNNQKWCYKDEINCKLEQNTIINN 350
OY 181 PKIISFNFELWYTYLLRDTIKNDKLCINNTTHCIDECKRNCLCFDRWYKQKEE 240
DB 351 PKIISFNFELWYTYLLRDTIKNDKLCINNTTHCIDECKRNCLCFDRWYKQKEE 410
OY 241 WNSIKKLFYTKKNNIQOSYSINNLFEGYFFKVDKLDKDEAKKELMENIKRKNFSN 300
DB 411 WNSIKKLFYTKKNNIQOSYSINNLFEGYFFKVDKLDKDEAKKELMENIKRKNFSN 470
OY 301 LENNRDYLENAIELLDHLEKETAT 324
DB 471 LENNRDYLENAIELLDHLEKETAT 494

RESULT 11

ID 08T6K8 PRELIMINARY; PRT; 496 AA.

AC 08T6K8;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
GN Erythrocyte membrane protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN NCBI_TaxID=5833;
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implied in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL: AF458896; AAL84271.1; -.
FT NON_TER 1 1
FT NON_TER 496 496
SQ SEQUENCE 496 AA; 59269 MW; A9569625061F9411 CRC64;

Query Match 83.6%; Score 1603; DB 5; Length 496;
Best Local Similarity 87.7%; Pred. No. 1, 2e-89;

Matches 286; Conservative 16; Mismatches 24; Indels 0; Gaps 0;

OY 1 YEKEIOSYLSNDKFNANNSEYKQFEKLEQYATNDFLNLNEGKCKGLPGEK 60
DB 171 YEKEIOSYLSNDKFNANNSEYKQFEKLEQYATNDFLNLNEGKCKGLPGEK 230
OY 61 DIFTNSADKGFYRSEYQVCPDGVKCDGKITYHKSNDREVRNEDYKPPWYKPT 120
DB 231 DIFTNSADKGFYRSEYQVCPDGVKCDGKITYHKSNDREVRNEDYKPPWYKPT 290
OY 121 NITVLYSGNQGDIQKLEFNCNSSTYKDKNNQKWCYKDEINCKLEQNTIINN 180
DB 291 NITVLYSGNQGDIQKLEFNCNSSTYKDKNNQKWCYKDEINCKLEQNTIINN 350
OY 181 PKIISFNFELWYTYLLRDTIKNDKLCINNTTHCIDECKRNCLCFDRWYKQKEE 240
DB 351 PKIISFNFELWYTYLLRDTIKNDKLCINNTTHCIDECKRNCLCFDRWYKQKEE 410
OY 241 WNSIKKLFYTKKNNIQOSYSINNLFEGYFFKVDKLDKDEAKKELMENIKRKNFSN 300
DB 411 WNSIKKLFYTKKNNIQOSYSINNLFEGYFFKVDKLDKDEAKKELMENIKRKNFSN 470
OY 301 LENNRDYLENAIELLDHLEKETAT 326
DB 471 LENNRDYLENAIELLDHLEKETAT 496

RESULT 12

ID 08T6K5 PRELIMINARY; PRT; 494 AA.

AC 08T6K5;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, last annotation update)
GN Erythrocyte membrane protein 1 (Fragment).
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
RN NCBI_TaxID=5833;
RP SEQUENCE FROM N.A.
RC STRAIN=K117;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implied in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL: AF458891; AAL84274.1; -.
FT NON_TER 1 1
FT NON_TER 494 494
SQ SEQUENCE 494 AA; 58870 MW; 56A34264A9285321 CRC64;

Query Match 82.2%; Score 1577; DB 5; Length 494;
Best Local Similarity 87.6%; Pred. No. 4, 5e-88;
Matches 283; Conservative 14; Mismatches 26; Indels 0; Gaps 0;

OY 1 YEKEIOSYLSNDKFNANNSEYKQFEKLEQYATNDFLNLNEGKCKGLPGEK 60
DB 172 YEKEIOSYLSNDKFNANNSEYKQFEKLEQYATNDFLNLNEGKCKGLPGEK 231
OY 61 DIFTNSADKGFYRSEYQVCPDGVKCDGKITYHKSNDREVRNEDYKPPWYKPT 120
DB 232 DIFTNSADKGFYRSEYQVCPDGVKCDGKITYHKSNDREVRNEDYKPPWYKPT 291
OY 121 NITVLYSGNQGDIQKLEFNCNSSTYKDKNNQKWCYKDEINCKLEQNTIINN 180
DB 292 NITVLYSGNQGDIQKLEFNCNSSTYKDKNNQKWCYKDEINCKLEQNTIINN 351
OY 181 PKIISFNFELWYTYLLRDTIKNDKLCINNTTHCIDECKRNCLCFDRWYKQKEE 240
DB 352 PKIISFNFELWYTYLLRDTIKNDKLCINNTTHCIDECKRNCLCFDRWYKQKEE 411
OY 241 WNSIKKLFYTKKNNIQOSYSINNLFEGYFFKVDKLDKDEAKKELMENIKRKNFSN 300
DB 471 WNSIKKLFYTKKNNIQOSYSINNLFEGYFFKVDKLDKDEAKKELMENIKRKNFSN 470

DB 412 WNSIKKLFYTKNNVPOPIYNNINLFEQYFFKVDKLDKNEAKKWEKMEINIKKKSEPSN 471
 QY 301 LNNRDYLENAIELLDLHMKETA 323
 DB 472 LNNRDYLENAIELLDLHMKETA 494

RESULT 13

Q876K1

ID 0876K1

PRELIMINARY; PRT; 460 AA.

AC 0876K1; 01-JUN-2002 (TREMblrel. 21, Created)

DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)

DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)

GN Erythrocyte membrane protein 1 (Fragment).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

SEQUENCE FROM N.A.

STRAIN=VAN002;

RT "Identification of a conserved Plasmodium falciparum var gene

RL J. Infect. Dis. 0:0-0(2002)."

DR EMBL; AF458895; AAL84278.1; -.

FT NON_TER 1 460

SQ SEQUENCE 460 AA; 55157 MW; 06FC87959F5C7B9E CRC64;

Query Match Best Local Similarity 76.4%; Score 1465; DB 5; Length 460;

Matches 260; Conservative 13; Mismatches 22; Indels 0; Gaps 0;

QY 1 YEKEIOSYLSNDKFNVNINSEYKOFYEKLETOYATNDTFLNLNEGKYGKGLPGEK 60

DB 165 YEKEIOSYLSNDKFNVNINSEYKOFHQLRDKNKNKNDTFLNLNEGKYGKGLPGEK 224

QY 61 DITFTNSADDKGIFRSEYQVCPDGVKCDGKIKYTHKSDNDRERNNDYKPPMGVKEPT 120

DB 225 DITFTNSADDKGIFRSEYQVCPDGVKCDGKIKYTHKSDNDRERNNDYKPPMGVKEPT 284

QY 121 NITVLSGNEOGDITOKLEFNCSSSTNYKDKNNQKWECKYKENTIRCKLEQNTETINDN 180

DB 285 NITVLSGNEOGDITOKLEFNCSSSTNYKDKNNQKWECKYKENTIRCKLEQNTETINDN 344

QY 181 PRTISFHFELWVYLLDRTIKMNDKLTCTCINNTTTHCIDEQNRNCLCFDRWVKOKEE 240

DB 345 PRTISFHFELWVYLLDRTIKMNDKLTCTCINNTTTHCIDEQNRNCLCFDRWVKOKEE 404

QY 241 WNSIKKLFYTKNNVPOPIYNNINLFEQYFFKVDKLDKNEAKKWEKMEINIKKK 295

DB 405 WNSIKKLFYTKNNVPOPIYNNINLFEQYFFKVDKLDKNEAKKWEKMEINIKKK 459

RESULT 14

Q876K2

ID 0876K2

PRELIMINARY; PRT; 427 AA.

AC 0876K2; 01-JUN-2002 (TREMblrel. 21, Created)

DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)

DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)

GN Erythrocyte membrane protein 1 (Fragment).

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

SEQUENCE FROM N.A.

STRAIN=TMAC8;

RT "Identification of a conserved Plasmodium falciparum var gene

RT "Identified in malaria in pregnancy."

RL J. Infect. Dis. 0:0-0(2002).

DR EMBL; AF458894; AAL84277.1; -.

FT NON_TER 1 427

SQ SEQUENCE 427 AA; 50969 MW; 6AF7922CB9B88F20 CRC64;

Query Match Best Local Similarity 65.7%; Score 1260; DB 5; Length 427;

Matches 222; Conservative 12; Mismatches 24; Indels 0; Gaps 0;

QY 1 YEKEIOSYLSNDKFNVNINSEYKOFYEKLETOYATNDTFLNLNEGKYGKGLPGEK 60

DB 170 YEKEIOSYLSNDKFNVNINSEYKOFYRRDKNKNKNDTFLNLNEGKYGKGLPGEK 229

QY 61 DITFTNSADDKGIFRSEYQVCPDGVKCDGKIKYTHKSDNDRERNNDYKPPMGVKEPT 120

DB 230 DITFTNSADDKGIFRSEYQVCPDGVKCDGKIKYTHKSDNDRERNNDYKPPMGVKEPT 289

QY 121 NITVLSGNEOGDITOKLEFNCSSSTNYKDKNNQKWECKYKENTIRCKLEQNTETINDN 180

DB 290 NITVLSGNEOGDITOKLEFNCSSSTNYKDKNNQKWECKYKENTIRCKLEQNTETINDN 349

QY 181 PRTISFHFELWVYLLDRTIKMNDKLTCTCINNTTTHCIDEQNRNCLCFDRWVKOKEE 240

DB 350 PRTISFHFELWVYLLDRTIKMNDKLTCTCINNTTTHCIDEQNRNCLCFDRWVKOKEE 409

QY 241 WNSIKKLFYTKNNVPOPIYNNINLFEQYFFKVDKLDKNEAKKWEKMEINIKKK 295

DB 410 WNSIKKLFYTKNNVPOPIYNNINLFEQYFFKVDKLDKNEAKKWEKMEINIKKK 427

RESULT 15

Q26032

ID 026032

PRELIMINARY; PRT; 3006 AA.

AC 026032; 01-NOV-1996 (TREMblrel. 01, Created)

DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)

DE 01-DEC-2001 (TREMblrel. 19, Last annotation update)

GN Variant-specific surface protein.

OS Plasmodium falciparum.

OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

SEQUENCE FROM N.A.

STRAIN=FCR3;

MEDLINE=95330813; PubMed=760678;

Su X.Z., Heacock V.M., Wertheimer S.P., Guinet F., Herrfeldt J.A.,

Peterson D.S., Ravetch J.A., Wellens T.E.;

"The large diverse gene family var encodes proteins involved in

cytoadherence and antigenic variation of Plasmodium falciparum-

infected erythrocytes."

RL Cell 82:89-100(1995).

DR EMBL; L40609; AAA75397.1; -.

DR InterPro; IPR004256; PfEMP.

DR Pfam; PF03011; PfEMP; 2.

SQ SEQUENCE 3006 AA; 343769 MW; 8FD5F475F943C74 CRC64;

Query Match Best Local Similarity 40.2%; Score 771; DB 5; Length 3006;

Matches 147; Conservative 72; Mismatches 115; Indels 20; Gaps 6;

QY 1 YEKEIOSYLSNDKFNVNINSEYKOFYEKLETOYATNDTFLNLNEGKYGKGLPGEK 60

DB 392 YTELEIKYSSSDKSSNSISKYNNEFYENNGKREYELQNFLLKNGATCOKEIEE 451

QY 61 DITFTNSADDKGIFRSEYQVCPDGVKCDGKIKYTHKSDNDRER-----VNEDYKP 113

DB 452 VIDE--NKDEDMVFRHSYCPDGVKCDGKIKYTHKSDNDRER-----VNEDYKP 506

QY 114 PWGKPTFTVLYSGNODGDTOKLEFNCSSSTNYKDKNNQKWECKYKENTIRCKLEQNT 173

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:25:50 ; Search time 3.48066 Seconds

(without alignments)
2916.372 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1918

Sequence: 1 YKEKIQSYLSNDNKFVNIN.....CRDNTNACETSHATTNP 345

Scoring table: BLOSUM62
Gapop. 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PTCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfilest1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	394.5	20.6	2710	2	US-08-568-459A-12
2	394.5	20.6	2710	2	US-08-487-826B-12
3	394.5	20.6	2710	4	US-09-210-288-12
4	394.5	20.6	3060	2	US-08-487-826B-14
5	380	19.8	700	2	US-08-568-459A-10
6	380	19.8	700	2	US-08-487-826B-10
7	380	19.8	700	4	US-09-210-288-10
8	380	19.8	2182	2	US-08-487-826B-16
9	231	12.0	921	2	US-08-568-459A-8
10	231	12.0	921	2	US-08-487-826B-8
11	231	12.0	921	2	US-09-210-288-8
12	145	7.6	749	2	US-08-568-459A-6
13	145	7.6	749	2	US-08-487-826B-6
14	145	7.6	749	4	US-09-210-288-6
15	137	7.1	1435	2	US-08-568-459A-4
16	137	7.1	1435	2	US-08-487-826B-4
17	137	7.1	1435	4	US-09-210-288-4
18	136	7.1	1147	3	US-08-470-260-5
19	136	7.1	1147	3	US-08-471-491-5
20	136	7.1	1147	4	US-08-466-662-5
21	136	7.1	3289	2	US-08-477-451-2
22	133	6.4	2391	2	US-08-446-855A-2
23	123	6.3	2391	4	US-09-150-741-2
24	120	6.3	463	4	US-08-845-258-25
25	120	6.3	463	4	US-08-990-571-25
26	120	6.3	463	4	US-08-723-142A-25
27	120	6.3	463	4	US-09-528-784A-25

28	118.5	6.2	1085	1	US-08-431-080-28	Sequence 28, Appl
29	118.5	6.2	1085	2	US-08-938-534-28	Sequence 28, Appl
30	118.5	6.2	1085	4	US-09-345-294-28	Sequence 28, Appl
31	118	6.2	1115	2	US-08-568-459A-2	Sequence 2, Appl1
32	118	6.2	1115	2	US-08-487-826B-2	Sequence 2, Appl1
33	118	6.2	1115	4	US-09-210-288-2	Sequence 2, Appl1
34	118	6.2	1115	6	5198347-6	Patent No. 5198347
35	116.5	6.1	290	2	US-08-903-801-1	Sequence 1, Appl1
36	116.5	6.1	290	4	US-09-295-055-1	Sequence 1, Appl1
37	114.5	5.9	1104	4	US-08-923-992A-4	Sequence 4, Appl1
38	112.5	5.9	1151	4	US-09-134-001C-3242	Sequence 3242, Ap
39	112	5.8	10182	4	US-09-134-001C-3159	Sequence 3159, Ap
40	110.5	5.8	1164	4	US-08-923-992A-2	Sequence 2, Appl1
41	110.5	5.8	1164	4	US-08-923-992A-10	Sequence 10, Appl
42	110	5.7	778	6	5198347-4	Patent No. 5198347
43	108	5.6	3135	1	US-08-323-170B-2	Sequence 2, Appl1
44	108	5.6	3135	4	US-08-954-441-2	Sequence 2, Appl1
45	107.5	5.6	599	2	US-08-910-551B-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-568-459A-12
Sequence 12, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim I.
APPLICANT: Chituls, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-12
Query Match 20.6% ; Score 394.5 ; DB 2 ; Length 2710 ;

Best Local Similarity 26.2%; Pred. No. 4.4e-25;
Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

```
QY 1 YEKELQSY-----LSNDRKFNVINNS-YYKQFYERKLEQYATNDFNLNLECK 50
Db 396 YDEIKRYENGASGSRQKRDAGGTTTNYDGEKKFYELNKSEYRTVDKLEKLSNEE 455
QY 51 YC-----KGLPGEKIDITFTNSA-----DDKGIIFYSEYQVCPDCGVKCDGIRKYTH 97
Db 456 ICTKYKDEEGGTTDFKNNVNSDSTSGASGTNVESQGFYRSKYCOPCYGVK-----KVNN 511
QY 98 -KSDNDRERNVN-----EDYKPPMGVKTPTNITVLYSGNEQGDITQKLENGCN-----143
Db 512 GGSNEMERKNGKCKSGKLYEPKPDKEGTTITLKSQGHDDIEKLNKFCDEKNGDTI 571
QY 144 -----SSTNYKDKNN-----QKMECYKDNENINCKLEONTLENNDPK-----11--184
Db 572 NSGSGTGGSGGSGRQELYEEMKC-YKGEDYKVVGHDEDEDEYENKNNAGGLCILKN 630
QY 185 -----SFHNFELWVYLLRDTIKWMDKLTCTI-NNTTTHC-ID 221
Db 631 QKKNKEGNTSEKPEDEIQKTENPFYVVAHMLKDSIMWKKLQRCLONGNRIRKCGNN 690
QY 222 EGNRNCIFDPMVKQKEEEMNSIKLFTKKNQ-----QSYYSNINLFE 267
Db 691 KGNNDCECFKRWITQKDEMGKIYQH-KTONIKRGSGDNTAELIPFDHYLYQYVLOE 749
QY 268 GYF-----FKVMDKLDKDEAKWELMENIKRKNKFSNLENNRD-----YLEN 310
Db 750 EFLKGDSEDAESEKSENSIDAEEA-----ELKHLREIIESEDNNOASVGGVTEQKN 803
QY 311 AIELLDHLKETATTC-----KONNTEACETSHNATNP 345
Db 804 IMDKLINTEKDLADCLIEHDEBEERKGDNECIEGENFRYNP 849
```

RESULT 2

US-08-487-826B-12
Sequence 12, Application US/08487826B
Patent No. 5993827

GENERAL INFORMATION:

APPLICANT: Slim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelisen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176

INFORMATION FOR SEQ. ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-12

Query Match

20.6%; Score 394.5; DB 2; Length 2710;
Best Local Similarity 26.2%; Pred. No. 4.4e-25;
Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

```
QY 1 YEKELQSY-----LSNDRKFNVINNS-YYKQFYERKLEQYATNDFNLNLECK 50
Db 396 YDEIKRYENGASGSRQKRDAGGTTTNYDGEKKFYELNKSEYRTVDKLEKLSNEE 455
QY 51 YC-----KGLPGEKIDITFTNSA-----DDKGIIFYSEYQVCPDCGVKCDGIRKYTH 97
Db 456 ICTKYKDEEGGTTDFKNNVNSDSTSGASGTNVESQGFYRSKYCOPCYGVK-----KVNN 511
QY 98 -KSDNDRERNVN-----EDYKPPMGVKTPTNITVLYSGNEQGDITQKLENGCN-----143
Db 512 GGSNEMERKNGKCKSGKLYEPKPDKEGTTITLKSQGHDDIEKLNKFCDEKNGDTI 571
QY 144 -----SSTNYKDKNN-----QKMECYKDNENINCKLEONTLENNDPK-----11--184
Db 572 NSGSGTGGSGGSGRQELYEEMKC-YKGEDYKVVGHDEDEDEYENKNNAGGLCILKN 630
QY 185 -----SFHNFELWVYLLRDTIKWMDKLTCTI-NNTTTHC-ID 221
Db 631 QKKNKEGNTSEKPEDEIQKTENPFYVVAHMLKDSIMWKKLQRCLONGNRIRKCGNN 690
QY 222 EGNRNCIFDPMVKQKEEEMNSIKLFTKKNQ-----QSYYSNINLFE 267
Db 691 KGNNDCECFKRWITQKDEMGKIYQH-KTONIKRGSGDNTAELIPFDHYLYQYVLOE 749
QY 268 GYF-----FKVMDKLDKDEAKWELMENIKRKNKFSNLENNRD-----YLEN 310
Db 750 EFLKGDSEDAESEKSENSIDAEEA-----ELKHLREIIESEDNNOASVGGVTEQKN 803
QY 311 AIELLDHLKETATTC-----KONNTEACETSHNATNP 345
Db 804 IMDKLINTEKDLADCLIEHDEBEERKGDNECIEGENFRYNP 849
```

RESULT 3

US-09-210-288-12
Sequence 12, Application US/09210288
Patent No. 6392026

GENERAL INFORMATION:

APPLICANT: Slim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible


```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1EMDVI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-12

Query Match      20.6%; Score 394.5; DB 4; Length 2710;
Best Local Similarity 26.2%; Pred. No. 4.4e-25;
Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

QY 1 YKEKIOSY-----LSNDKRVNNINSE-YKQFEYKLEQVATNDFLLMLEGK 50
D 396 YDEIKRYENGASGSGKQKRDAGGTTTNDYGEKKFYDELNKSSEYTVKFEKLSNEE 455
QY 51 YC-----KGLPGEKIDITFTNSA-----DQKGIYRSEYCOVPCDGCYKCGIRYTH 97
D 456 ICTKVKDEEGTIDFKVNVSDSTSGASGTNYESGTFYRSKTCQPCYCGVK-----KVN 511
QY 98 -KSDNDERVNN-----EDYKPPWGVKPTNITVLYSGNEGDTTOKLENFCA-----143
D 512 GGSSENEWEKRNKCKSGKLYEPKDEGTTITLKSGKHGHDIEKLNKFCDEKNGDTI 571
QY 144 -----SSTNYDKNN-----QKWECCYKDEINRCKLEONTLEINNDPK-----11-- 184
D 572 NSGSGTGGSGGSGSGQELYEEMKC-YKGEDVYKVGHDDEDEDEYEVKNAAGLCITLKN 630
QY 185 -----SFTNFEELWVYLLRDTIKMNDKLTCTI--NNTTTHC-ID 221
D 631 QKKNKEGSGTSEKPEIDQKTFNPFYVVAHMLKDSIMKKKLQKLCLONGNRIKCGNN 690
QY 222 ECNRNCLCFDRWVYQKEEENSIKKLFTKKNNIO-----QSTYSINNLF 267
D 691 KGNNDCECFKRWITQKDEMGKIYQHF-KTONIGRSGSDNTAELLIPDHDYVLQYNIQ 749
QY 268 GYF-----FKVMOKLQDEAKWELMENIKRKNNEFNLNNND-----YLEN 310
D 750 EFLKGDSEDASEESENSLDAEFA-----ELKHLRELTIESEDNNOEASVGGVGTBOKN 803
QY 311 AIELLDHLKETATIC-----KDNNTNEACETSHATTNP 345
D 804 IMOKLWYKEDADLCLEIHDEDEEKEKGDGNCIEGGEFRINP 849

RESULT 4
US-08-487-826B-14
Sequence 14, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chiltons, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su Xian-zhaun
APPLICANT: Wellens, Thomas E.

```

```

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 3060 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-487-826B-14

```

```

Query Match      20.6%; Score 394.5; DB 2; Length 3060;
Best Local Similarity 26.2%; Pred. No. 5.2e-25;
Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

QY 1 YKEKIOSY-----LSNDKRVNNINSE-YKQFEYKLEQVATNDFLLMLEGK 50
D 394 YDEIKRYENGASGSGKQKRDAGGTTTNDYGEKKFYDELNKSSEYTVKFEKLSNEE 453
QY 51 YC-----KGLPGEKIDITFTNSA-----DQKGIYRSEYCOVPCDGCYKCGIRYTH 97
D 454 ICTKVKDEEGTIDFKVNVSDSTSGASGTNYESGTFYRSKTCQPCYCGVK-----KVN 509
QY 98 -KSDNDERVNN-----EDYKPPWGVKPTNITVLYSGNEGDTTOKLENFCA-----143
D 510 GGSSENEWEKRNKCKSGKLYEPKDEGTTITLKSGKHGHDIEKLNKFCDEKNGDTI 569
QY 144 -----SSTNYDKNN-----QKWECCYKDEINRCKLEONTLEINNDPK-----11-- 184
D 570 NSGSGTGGSGGSGSGQELYEEMKC-YKGEDVYKVGHDDEDEDEYEVKNAAGLCITLKN 628
QY 185 -----SFTNFEELWVYLLRDTIKMNDKLTCTI--NNTTTHC-ID 221
D 629 QKKNKEGSGTSEKPEIDQKTFNPFYVVAHMLKDSIMKKKLQKLCLONGNRIKCGNN 688
QY 222 ECNRNCLCFDRWVYQKEEENSIKKLFTKKNNIO-----QSTYSINNLF 267
D 689 KGNNDCECFKRWITQKDEMGKIYQHF-KTONIGRSGSDNTAELLIPDHDYVLQYNIQ 747
QY 268 GYF-----FKVMOKLQDEAKWELMENIKRKNNEFNLNNND-----YLEN 310
D 748 EFLKGDSEDASEESENSLDAEFA-----ELKHLRELTIESEDNNOEASVGGVGTBOKN 801
QY 311 AIELLDHLKETATIC-----KDNNTNEACETSHATTNP 345
D 802 IMOKLWYKEDADLCLEIHDEDEEKEKGDGNCIEGGEFRINP 847

RESULT 5

```

US-08-568-459A-10
; Sequence 10, Application US/08568459A
; Patent No. 5849306
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/568,459A
; FILING DATE: 07-DEC-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelien, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH21.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-568-459A-10

Query Match 19.8%; Score 380; DB 2; Length 700;
Best Local Similarity 28.7%; Pred. No. 1.3e-24;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

QY 1 YEKET-----QSYLSDNKFVNNINSEYKOFYKLETOYAINDTFL 43
DB 285 YETELISGGSGSPKRTKRAARSSSSDDN-----GYESKFYKLEKVEGYQVDVDFL 336
QY 44 NLNEGKCYCG-----GLPGEKDIPTNSADDKGIFRSEYCVCPGCGYKCGIKYTHS 99
DB 337 KLINKEGICQKOPQYGNKADNDVFTNEKYK-TESRTEICEPCWCGLEKGGPPKVKYG 395
QY 100 DNDRRERNEDYKPPGVKFTNITVLYSGNEGDTITQKLENCNSSTNYKDKNNOKWECY 159
DB 396 DKTGSATKTYDPK---NTIDIPVLYPKSQONILKRYKNCEKAP-GGGQIKKMGY 451
QY 160 YKD-----ENINRCKLEONTEINNDNPKIISPHNFEFLAVYLLADDTIKWMDKLTCTINN 214
DB 452 YDEHPPSSKNNNNVCGTMDKFTQKQVYKYNVFPWVMDLHDVEMTELSKCIINN 511
QY 215 TTT-----HCIDECNRNCICFDRAWYKQKEEENSJIKLFTKKNNIQOASYYSININLFESEY 270
DB 512 NTNGTCNRNNKCKTKDCCCFQKWEKKQOEWMAIKDHFGKQTDIVQO-----KGLIYFSP 566
QY 271 FKVMDKLDKDEAKWELMENIKRKKNESENENNRDYLE-----NAIELL 316

US-08-487-826B-10
; Sequence 10, Application US/08487826B
; Patent No. 5993827
; GENERAL INFORMATION:
; APPLICANT: Sim, Kim L.
; APPLICANT: Chitnis, Chetan
; APPLICANT: Miller, Louis H.
; APPLICANT: Peterson, David S.
; APPLICANT: Su, Xin-zhaun
; APPLICANT: Wellens, Thomas E.
; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
; TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,826B
; FILING DATE: 10-SEP-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelien, Ned
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH21.001CPI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 700 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
US-08-487-826B-10

Query Match 19.8%; Score 380; DB 2; Length 700;
Best Local Similarity 28.7%; Pred. No. 1.3e-24;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

QY 1 YEKET-----QSYLSDNKFVNNINSEYKOFYKLETOYAINDTFL 43
DB 285 YETELISGGSGSPKRTKRAARSSSSDDN-----GYESKFYKLEKVEGYQVDVDFL 336
QY 44 NLNEGKCYCG-----GLPGEKDIPTNSADDKGIFRSEYCVCPGCGYKCGIKYTHS 99
DB 337 KLINKEGICQKOPQYGNKADNDVFTNEKYK-TESRTEICEPCWCGLEKGGPPKVKYG 395
QY 100 DNDRRERNEDYKPPGVKFTNITVLYSGNEGDTITQKLENCNSSTNYKDKNNOKWECY 159
DB 396 DKTGSATKTYDPK---NTIDIPVLYPKSQONILKRYKNCEKAP-GGGQIKKMGY 451

QY 160 YKD-----ENINCKLEONTNNDNPKIISFHFPELWYTLRDTIKNDKLTCTINN 214
 Db 452 YDEHRSSKNNNNCVBETWMTKFTOGKOTVKSYNFFWDMVHDLHDSVEKTELKSCINN 511
 QY 215 TTT-----HIDEENRNCICFDRAWYKQKEEEMNSIKLFTKKNKIQOSYSININLFEYGF 270
 Db 512 NTNGNTRNNKCKTDCGCKQKWEKKQOEMAIKIHFGKOTDIVQO-----KGLIYFSP 566
 QY 271 FKVDKDKDEAKWKEEMENIKRKNKNEFSNLNNDYLE-----NAIEELL 316
 Db 567 YGVLDLV-----LKGNNLQNIKDVHGDTDDIKHKKLLDEDAVAVVVGKNDTTIDKLL 622
 QY 317 DHIKETATICKDNNTNEACE 336
 Db 623 QHEKEQAEOCKOK--OECE 640

RESULT 7
 US-09-210-288-10
 Sequence 10, Application US/09210288
 Patent No. 6392026
 GENERAL INFORMATION:
 APPLICANT: Slim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/210, 288
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Fuller, Michael
 REGISTRATION NUMBER: 36,516
 REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 10:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 700 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ORIGINAL SOURCE:
 ORGANISM: Plasmodium falciparum
 US-09-210-288-10

Query Match 19.8%; Score 380; DB 4; Length 700;
 Best Local Similarity 28.7%; Pred. No. 1.3e-24;
 Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

QY 1 YEKEI-----OSYLSNDKRFVNNINSEYKQFERKLETOYATNDTFL 43
 Db 285 YETELSGGSGKSPKRTKRAARSSSSSDN-----GYESKRYKKLKEGYDVVKFL 336

QY 44 NLNEGKYCKG----GLPGEKDTITNSADDKGIFRSEYQVCPDGVKCDGIKYTHKS 99
 Db 337 KILNKEGICQKPOVGNKEKADNDVFTNEKYK--FESRTEICEPCPWCLEKGGPMMVKG 395
 QY 100 DNDREKRNNDYKPPMKVKTNTIVLYSGNQGDTOTKLENFCSSTNYKDKNNQKECY 159
 Db 396 DKTCGSAKTYTDEK---NTIDIPVLVDPKSOQNLKRYKPFCEKGP--GGQIKRWQCY 451
 QY 160 YKD-----ENINCKLEONTNNDNPKIISFHFPELWYTLRDTIKNDKLTCTINN 214
 Db 452 YDEHRSSKNNNNCVBETWMTKFTOGKOTVKSYNFFWDMVHDLHDSVEKTELKSCINN 511
 QY 215 TTT-----HIDEENRNCICFDRAWYKQKEEEMNSIKLFTKKNKIQOSYSININLFEYGF 270
 Db 512 NTNGNTRNNKCKTDCGCKQKWEKKQOEMAIKIHFGKOTDIVQO-----KGLIYFSP 566
 QY 271 FKVDKDKDEAKWKEEMENIKRKNKNEFSNLNNDYLE-----NAIEELL 316
 Db 567 YGVLDLV-----LKGNNLQNIKDVHGDTDDIKHKKLLDEDAVAVVVGKNDTTIDKLL 622
 QY 317 DHIKETATICKDNNTNEACE 336
 Db 623 QHEKEQAEOCKOK--OECE 640

RESULT 8
 US-08-487-826B-16
 Sequence 16, Application US/08487826B
 Patent No. 5993827
 GENERAL INFORMATION:
 APPLICANT: Slim, Kim L.
 APPLICANT: Chitnis, Chetan
 APPLICANT: Miller, Louis H.
 APPLICANT: Peterson, David S.
 APPLICANT: Su, Xin-zhaun
 APPLICANT: Wellens, Thomas E.
 TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
 TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
 NUMBER OF SEQUENCES: 45
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe Martens Olson & Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: California
 COUNTRY: US
 ZIP: 92660
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/487, 826B
 FILING DATE: 10-SEP-1993
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Israelson, Ned
 REGISTRATION NUMBER: 29,655
 REFERENCE/DOCKET NUMBER: NIH121.001CPI
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (619) 235-8550
 TELEFAX: (619) 235-0176
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2182 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 FRAGMENT TYPE: N-terminal

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelsen, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-487-826B-8

Query Match 12.0%; Score 231; DB 2; Length 921;
Best Local Similarity 24.1%; Pred. No. 1,3e-11;
Matches 91; Conservative 55; Mismatches 147; Indels 84; Gaps 17;

QY 36 YATNDTFLNLNBSKYKGLPGKEDITFTNSADKGIIFYRSEYQYCP----DCGYKCD 91
DB 6 YTSIGLFLNLSKSGPCCKNDNAEDNIDF---GDEGTFKFEADCKRCQSFYVDC-KNCN 61
QY 92 GIKYTHKSDNDREVRNEDYKPPWGVKPTNITVLYSGNEGDTIOKLE-----NF 141
DB 62 GGDYKRGKNGSNGKKNNDY-----ITASDIENGNSIGNIDMVSDKANGFNGLDA 114
QY 142 CNSSTYKDKNNQWECYVDENINRCKLEQ-NTEINNDNPKIISFNFEFLWYTYLLRD 200
DB 115 CGSANIKFGIRKEQWKC-AKYCGLDVGLKNGSIDKQKQIITIRALKRWVEYLED 173
QY 201 TIKWMDLKTICI--NNTTHCIDECCNRNCLCFDRWVQKEEWSIKKLF-TKKNNIOOS 257
DB 174 YKNTNAKISHCTKKNDEST-CTNDCPNKCTCVEEMINQKTEWKNIKKHKTQENGDNN 232
QY 258 YYSNINNL-----FEGYFFVMDKLDKDEAKWKE-----LM 288
DB 233 MSLVTDILGALQPSDVNKAIRKPCSLTAPES--FCGLNADNSEKKEGEDYDVLVCLM 290
QY 289 ENIK-----RKNEFSNLEN-----NRDYLENAIELLDHLKETATIC-----KD 328
DB 291 KMLEKQIOECKKKHGETSVENGSKSTPLDNTLTLEEPPIE--EENQVEAPNICPKQTVED 348
QY 329 NNTNEACETSHNATTNP 345
DB 349 KKKEEETCTPASVP 365

RESULT 11
US-09-210-288-8
Sequence 8, Application US/09210288
Patent No. 6392026
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM EXYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US

ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 921 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-8

Query Match 12.0%; Score 231; DB 4; Length 921;
Best Local Similarity 24.1%; Pred. No. 1,3e-11;
Matches 91; Conservative 55; Mismatches 147; Indels 84; Gaps 17;

QY 36 YATNDTFLNLNBSKYKGLPGKEDITFTNSADKGIIFYRSEYQYCP----DCGYKCD 91
DB 6 YTSIGLFLNLSKSGPCCKNDNAEDNIDF---GDEGTFKFEADCKRCQSFYVDC-KNCN 61
QY 92 GIKYTHKSDNDREVRNEDYKPPWGVKPTNITVLYSGNEGDTIOKLE-----NF 141
DB 62 GGDYKRGKNGSNGKKNNDY-----ITASDIENGNSIGNIDMVSDKANGFNGLDA 114
QY 142 CNSSTYKDKNNQWECYVDENINRCKLEQ-NTEINNDNPKIISFNFEFLWYTYLLRD 200
DB 115 CGSANIKFGIRKEQWKC-AKYCGLDVGLKNGSIDKQKQIITIRALKRWVEYLED 173
QY 201 TIKWMDLKTICI--NNTTHCIDECCNRNCLCFDRWVQKEEWSIKKLF-TKKNNIOOS 257
DB 174 YKNTNAKISHCTKKNDEST-CTNDCPNKCTCVEEMINQKTEWKNIKKHKTQENGDNN 232
QY 258 YYSNINNL-----FEGYFFVMDKLDKDEAKWKE-----LM 288
DB 233 MSLVTDILGALQPSDVNKAIRKPCSLTAPES--FCGLNADNSEKKEGEDYDVLVCLM 290
QY 289 ENIK-----RKNEFSNLEN-----NRDYLENAIELLDHLKETATIC-----KD 328
DB 291 KMLEKQIOECKKKHGETSVENGSKSTPLDNTLTLEEPPIE--EENQVEAPNICPKQTVED 348
QY 329 NNTNEACETSHNATTNP 345
DB 349 KKKEEETCTPASVP 365

RESULT 12
US-08-568-459A-6
Sequence 6, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.

APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FMDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 749 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-09-210-288-6

Query Match 7.6%; Score 145; DB 4; Length 749;
Best Local Similarity 21.5%; Pred. No. 0.00026;
Matches 61; Conservative 46; Mismatches 103; Indels 74; Gaps 11;

QY 91 DGIR-VTHKSDNDRERANNEDYKRPVGVKPTNITVLYSGNEOGDITQKLEFNCSSYNYK 149
D 162 EGLMDHINKANYEAMHLEKREYENAGDGKICN-AIIGSYADIGDIYVAGLDVMDINTN-- 218
D 150 DKNNAQKECYKKNINRCLEONTETNN-----DNPKTISHNFEEL 192
D 219 -KISEKQKLFMGGSNRKQNDNNERNKWKORNLWSSMVKHPRKGTCKRHNNFEK 277
QY 193 WVTYLLBDITKMDKL-----KTCINNTTHCIDEGRNCLCFDRWVKOKEE 240
D 278 -IQPLFMKLEWGDPECEBEGTEKQLEKICENKCE--KCCNACSSYETKMKKEKNE 334
QY 241 WNSIKLFTKKKNIQOQSYYSINNNLEGEYFEKVDKLDKDEAKKELMENIKRRKNEFSN 300
D 335 YNLSQSKFSDSKLNRK--NNLYN-----KREDSKAVLRSESKQCSN 374
QY 301 LNNRQYLENAIELLDHLKETATICDNNNTNACETSHNATN 344
D 375 IEFN-----DETFTF--PNKYKCAVCNPPSSS 401

RESULT 15
US-08-568-459A-4
Sequence 4, Application US/08568459A
Patent No. 5849306
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnais, Chetan
APPLICANT: Miller, Louis H.

APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhaun
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/568,459A
FILING DATE: 07-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CPI
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
US-08-568-459A-4

Query Match 7.1%; Score 137; DB 2; Length 1435;
Best Local Similarity 20.5%; Pred. No. 0.003;
Matches 86; Conservative 57; Mismatches 149; Indels 128; Gaps 19;

QY 1 YEKIQSYLSDN-----KFVN-NINSEYKOFYEKLEQYATNDLFNLNGKYCK- 53
D 75 FNRKRSYGPNDNDKMSLKNKNNEMFNNTQSLSTS-----SLIKONKIVPI 126
QY 54 GGLPEKIDITFTNSADKGLFYRSEYCOVCPDGVKCDGKITYHKSDNDRER-VNEDYK 112
D 127 NAVRVSILSLFSDRINNG-RNTSSNNEVLNSCKREKRGKMKWDCKKNDNSNYCIPDRR 185
QY 113 PRQVGVKPTNITVLYSGNEOGD-----ITQKLEN-----FCNSSYN----- 147
D 186 IQLCIVNLSTIKTYKTMKDHFTLASKRESQLLKKNNDKNSKFCNDLKNSELYGHL 245
QY 148 -----YKDKNNOK-----WECYKND--E 163
D 246 AMGNDMPGCGSTAEKNKIQVFAGCHGEISEHKIKFRKKNWMPREKMLSEAKN 305
QY 164 NINRCLEONTETNNDPKTIISFNFEELWVYLLRDTIKMDKLKTCINNTTH-CIDE 222
D 306 NINCKNIPQEL-QITQWIKEMHGEF-----LLERDNRSLPKSK-CKNNTLYEACEKE 358
QY 223 CNRNCLEFDRWVKOKEEEMNSIKLFTKKKNIQOQSYYSINNNLEGEYFEKVDKLDDEA 282
D 359 CIDPCKMYRQWITRSKEWHLISKEYETQKVPKN-----AENYLKIS--NNDA 408
QY 283 KWKELMENIKRRKNEFSNLENNRQYLENAIELLDHLKETATICK-----DNNNTNEACE 336
D 409 KVSLLANNCAEYKYCD-----CKHTTTLVKSVLNGNDNTIKERRE 450

Mon Apr 28 13:47:52 2003

us-10-087-013-2_copy_403_747.rai

Page 10

Search completed: April 28, 2003, 10:32:06
Job time : 9.48066 secs

2-1-0

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003. Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:32:15 ; Search time 7.70718 Seconds
(without alignments)
3586.886 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 1918

Sequence: 1 YKEQTQSYLSNDNKFVNIN.....CKDNTNEACETSHNATNP 345

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 301932 seqs, 80129803 residues

total number of hits satisfying chosen parameters: 301932

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	394.5	20.6	2710	9	US-10-153-273-12
2	380	19.8	700	9	US-10-153-273-10
3	231	12.0	921	9	US-10-153-273-8
4	162	8.4	807	9	US-09-820-843A-108
5	161	8.4	972	10	US-09-924-154-16
6	151.5	7.9	1086	10	US-09-924-154-15
7	147	7.7	1421	10	US-09-924-154-13
8	145	7.6	749	9	US-10-153-273-6
9	144	7.5	861	9	US-09-820-843A-109
10	137	7.1	1435	9	US-10-153-273-4
11	136	7.1	1336	10	US-09-402-100-4
12	133	6.9	1501	10	US-09-924-154-17
13	128	6.7	1143	10	US-09-924-154-14
14	122	6.4	1805	9	US-09-820-843A-73
15	120	6.3	463	9	US-09-286-488-25
16	120	6.3	463	10	US-09-737-178-25
17	118	6.2	996	10	US-09-815-242-5251
18	118	6.2	1009	10	US-09-815-242-12141
19	118	6.2	1115	9	US-10-153-273-2

20	116.5	6.1	290	10	US-09-808-885-1	Sequence 1, Appli
21	114	5.9	452	10	US-09-737-178-123	Sequence 123, App
22	107	5.6	1331	9	US-10-087-464-50	Sequence 50, Appl
23	106.5	5.6	652	10	US-09-351-794A-2	Sequence 2, Appli
24	106	5.5	521	9	US-09-820-843A-93	Sequence 93, Appl
25	104.5	5.4	858	9	US-10-011-588-41	Sequence 41, Appl
26	104.5	5.4	961	10	US-09-801-368-132	Sequence 132, App
27	104	5.4	2025	10	US-09-815-242-5703	Sequence 5703, Ap
28	104	5.4	3158	10	US-09-815-242-12611	Sequence 12611, A
29	102.5	5.3	665	9	US-09-820-843A-107	Sequence 107, App
30	100.5	5.2	394	9	US-09-978-756-3	Sequence 3, Appli
31	100.5	5.2	428	9	US-10-087-464-46	Sequence 46, Appl
32	99.5	5.2	383	9	US-10-098-514-2	Sequence 2, Appli
33	99.5	5.2	383	9	US-10-098-514-4	Sequence 4, Appli
34	99.5	5.2	402	9	US-10-098-514-14	Sequence 14, Appl
35	99.5	5.2	497	9	US-09-820-843A-32	Sequence 32, Appl
36	99	5.2	621	10	US-09-856-247A-2	Sequence 2, Appli
37	99	5.2	2789	10	US-09-801-574-57	Sequence 57, Appl
38	96.5	5.0	334	9	US-10-087-464-33	Sequence 33, Appl
39	96.5	5.0	360	9	US-10-087-464-12	Sequence 12, Appl
40	96.5	5.0	583	10	US-09-924-654-4	Sequence 4, Appli
41	96.5	5.0	704	10	US-09-801-368-218	Sequence 218, App
42	96.5	5.0	900	12	US-10-071-751-21	Sequence 21, Appl
43	96.5	5.0	1295	10	US-09-726-949A-1	Sequence 1, Appli
44	96.5	5.0	1639	9	US-10-087-464-10	Sequence 10, Appl
45	96	5.0	376	9	US-09-978-756-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-153-273-12

; Sequence 12, Application US/10153273

; Patent No. US20020169305A1

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; Chitnls, Chetan

; Miller, Louis H.

; Peterson, David S.

; Su, Xin-zhaun

; Wellemis, Thomas E.

; TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

; AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Knobbe Martens Olson & Bear

; STREET: 620 Newport Center Drive 16th Floor

; CITY: Newport Beach

; STATE: California

; COUNTRY: US

; ZIP: 92660

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/153,273

; FILING DATE: 21-May-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/210,288

; FILING DATE: <Unknown>

; ATTORNEY/AGENT INFORMATION:

; NAME: Fuller, Michael

; REGISTRATION NUMBER: 36,516

; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (619) 235-8550

; TELEFAX: (619) 235-0176

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

LENGTH: 2710 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 12;
US-10-153-273-12

Query Match 20.6%; Score 394.5; DB 9; Length 2710;
Best Local Similarity 26.2%; Pred. No. 7.7e-21;
Matches 122; Conservative 76; Mismatches 135; Indels 133; Gaps 20;

Qy 1 YKEIQSY-----LSNDNKFVNINSE-YKQFYKELKETQYATNDFLNLNGK 50 ~
Db 396 YDEIKKYENGASGSRQKRDAGTTTNYDQYKFKFDLKNKSYRIVDFLEKLSNEE 455
Qy 51 YC-----KGLPGKEDITFTNSA-----DDKGFYRSEYQVCPDCGVKCDGIKTYH 97
Db 456 ICTVKDERGGTIDFKNVNSDSTSGASGTVNBSQGTFRYSKYQPCPYCGVK-----KVNN 511
Qy 98 -KSDNDRVNN-----EDYPPGVKPTNITVLYSGNEQGDITQKLENECN-----143
Db 512 GGSSEWEKNGKCKGKLYEPKPDKEGTTITILKSGKHGDDIEKLNKFCDEKNGDTI 571
Qy 144 ----SSTNYKDKNN-----QKWCYKDNINRCKLEQNTNNDNPK-----II-- 184
Db 572 NSGGSTGGSGGNGRQELYEKWK-YKGEDVVKVGHDEDEEDYENVKNAGGLCILKN 630
Qy 185 -----SFHNFELWVYLLRDTIKWNDKLTCTI-NNTTHC-ID 221
Db 631 OKKNKEGGNTSEKPEDEIQTFFNFFYYWAHLKDSIHWKKLQRCIQNGNRKICGN 690
Qy 222 ECRNCLCFDRVVKQKEEWNISIKLFTKKKNTQ-----QSYYSNINNLFEE 267
Db 691 KCNDCCEPKRITOKKDWGKIVQHF-KTONIKRGSGSDNTAEILPDPHDVLYQYNQOE 749
Qy 268 GYF-----FKYMDKLDKDEAKWKMENIKRKKNEFSNLENRD-----YLEN 310
Db 750 EFLGDSSEDAESEKSENSLDAEEA-----BELKHLREIESEDNNOEASVGGVTEQKN 803
Qy 311 ATELLDHLKETATC-----KDNNTNEACETSHNATNP 345
Db 804 IMDKLLNYEKDADLCLETHEDEEEKEKGDGNECIEGENFRYP 849

US-10-153-273-10
Sequence 10, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.

TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NTH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 700 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 10;
US-10-153-273-10

Query Match 19.8%; Score 380; DB 9; Length 700;
Best Local Similarity 28.7%; Pred. No. 1.9e-20;
Matches 109; Conservative 55; Mismatches 148; Indels 68; Gaps 12;

Qy 1 YEKEI-----QSYLSNDNKFVNINSEYKQFYKELKETQYATNDFL 43
Db 285 YETEISGGSGKSPKRTKRAARSSSSDDN-----GYESKFKYKLEVGQYQDVDFL 336
Qy 44 NLLNEGKCKG---GLPEKDIFTNSADKGFYRSEYQVCPDCGVKCDGIKTYHKS 99
Db 337 KILNKEGICQKQVQGNKADNVDFTEKYK-TFSRTEICEPCWGLEKGGPPWKVG 395
Qy 100 DNDREVRVNEYDKPPWGVKPTNITVLYSGNEQGDITQKLENCNSSTNYKDKNNQWECY 159
Db 396 DKTCGSATKYDPK---NITDIPVLYPKSQONILKVKYKNCFKGAP-GGQIKKWQCY 451
Qy 160 YKD-----ENINRCKLEQNTNNDNPKLIISFHNFFELWVYLLRDTIKWNDKLTCTINN 214
Db 452 YDEHRPSSKNNNNVEGTWDFKFTQKQTVKSYNVFFWDMVHMLSDSVETKELSKCINN 511
Qy 215 TTT---HCIDECNRNCLCFDRVVKQKEEWNISIKLFTKKKNIQSYYSNINNLFEGYF 270
Db 512 NTNGYTCRNNKCKYDCGCFQKWVKKQOEWMIAIKDFGKQTDIVQQ-----KGLIVFSP 566
Qy 271 FKVMDKLDKDEAKWKMENIKRKKNEFSNLENNRDYLE-----NAIELLL 316
Db 567 YGVLDLV---LKGNNLLQNKIDVHGDITDDIKHKLLDEEDAVAVLGGKDNNTIDKLL 622
Qy 317 DHLKETATICKDNNTNEACE 336
Db 623 QHEKEQAEQCKQK--QEECE 640

RESULT 3
US-10-153-273-8
Sequence 8, Application US/10153273
Patent No. US20020169305A1
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
Chitnis, Chetan
Miller, Louis H.
Peterson, David S.
Su, Xin-zhaun
Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING PROTEINS

```

; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobe Martens Olson & Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: California
; COUNTRY: US
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/153,273
; FILING DATE: 21-May-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,288
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Fuller, Michael
; REGISTRATION NUMBER: 36,516
; REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 921 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium falciparum
; SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-10-153-273-8

Query Match 12.0%; Score 231; DB 9; Length 921;
Best Local Similarity 24.1%; Pred. No. 3 le-09;
Matches 91; Conservative 55; Mismatches 147; Indels 84; Gaps 17;

QY 36 YATNDTFLNLLNGKYCKGLPGEKDTFTNSADKGIYRSEYQVCP----DCGVKCD 91
DB 6 YTSIGLFLNSLKGPCCKDNDNAEDNIDF---GDEGTFKEADNCKPCSQFTVDC-KNCN 61
92 GIKYTHKSDNDRVRNVEDYKPPGWYKPTNITVLYSGNEQGDITQKLE-----NF 141
DB 62 GBDTKGKNGSGKNGNDY-----ITASDIENGSGNSIGNIDMVVSDKDANGFNGLDA 114
142 CNSSTNYKKNQKWEYKDEINCKLBO-NTEINNDNPKIISPHNFELWVYLLRD 200
DB 115 CGSANIFKGRKEQWK-AKVCGLDVCLKNGSIDKQKQIIIRALLKRWEYFLED 173
QY 201 TIKWNDKLKTCI--NNTTTHCIDECNRCICFDRWVKQKEEWNISIKLF-TRKKNIQOS 257
DB 174 YNKINAKISHCTKKNDEST-CTNDPCNKCVCVEWINKRTWKNIKKHYKTONENGDN 232
QY 258 YNSNNNL-----FEYFFKVMKDKDEAKWE-----LM 288
DB 233 MKSLVTDILGALQPSQDVNKAIRKPCSGLTAFES--FCGLNGADNSENKEGEDYDLVLCML 290
QY 289 ENIK-----RKNEFSNLEN-----NRDYLENAIELLDHLKETATC-----KD 328
DB 291 KNLKQIQEKKKHGKTSVNGSKSCTPLDNTTLEEEPIE--EENQVEAPNICPKQTVED 348
QY 329 NNTNEACETSHNATNP 345
DB 349 KKKEEBETCTPASVP 365

```

```

RESULT 4
US-09-820-843A-108
; Sequence 108, Application US/09820843A
; Publication No. US2003003963A1
; GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 108
; LENGTH: 807
; TYPE: PRT
; ORGANISM: Plasmodium falciparum
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: hypothetical protein
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi13845292
US-09-820-843A-108

Query Match 8.4%; Score 162; DB 9; Length 807;
Best Local Similarity 19.7%; Pred. No. 0.00035;
Matches 71; Conservative 55; Mismatches 118; Indels 116; Gaps 13;

QY 8 YLSNDKFNVINSEYKQFYELKKEQYA--TNDTFLNLLNGKYCKGLPGEKDTIF 64
DB 49 HMSNDS-----NINKQENKKKKKKTKKNNINNNTHINI----- 84
QY 65 TNSADDRG-----IFYRSEYQVCPDCGVKDCGIKTYTHKSDNDRVRNVEDYKPPW 115
DB 85 -HTNDKNGDINKPEVIERDNIINIKNDTNLIDSSYNEEGNENRNDINNNNNNI 143
QY 116 GVPRTNITVLYSGNEQGDITQKLEFCNSSTNYKKNQKWEYKDEINRCKLEQNT 175
DB 144 NINNNNI-----NNSCSNYYGLKKITLLKRNNDIKDEGYNN---ENITT 184
QY 176 INNDNPKIISPHNFELWVYLLRDTIKWNDKLKTCINNTTTHCIDECNRCICFDRWVK 235
DB 185 LNKNN-----LNKNYNDNRNN--NNNNKNNNNNNNNCCSEKLE 226
QY 236 QKEEWNISIK-KLFT----KKKNIQOSYYSINNLFEGYFFKVMKDKDEAKWKLME 290
DB 227 QREKEYNKIRARIFSNFKQKNVQKTEQNNLNHTYLN-----NNIINN 270
QY 291 IKRKNNEFSNLEN-----NRDYLENAIELLDHLKETATICKDNNTNEACETSHN 340
DB 271 INNGDNOYAYINNFIYHNNSYNNHYRONNI-----PICNINNHAPNIEKLNN 319

RESULT 5
US-09-924-154-16
; Sequence 16, Application US/09924154
; Patent No. US2002012741A1
; GENERAL INFORMATION:
; APPLICANT: Narum, David L.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Anti-plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; CURRENT FILING DATE: 2001-08-07
; PRIOR APPLICATION NUMBER: US 60/223,525
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 16
; LENGTH: 972
; TYPE: PRT
; ORGANISM: Mammalian

```

US-09-924-154-16

Query Match 8.4%; Score 161; DB 10; Length 972;

Best Local Similarity 22.7%; Pred. No. 0.00052;

Matches 77; Conservative 48; Mismatches 94; Indels 120; Gaps 21;

QY 9 LSNDRKFN-----INSEYKOFYKEL--KETOYATNDTFLNLNE-----48

Db 105 ICKNKNIALYKVEDICNNTKVSIGELYCKE---KGNCKIWCINIEHKIDFPDVC 161

QY 49 -----GKYCKGGLPGEKIDITFNSAD-----DKGIFRSEYC-----80

Db 162 GPPRRQQLCLGNL--DKD-EFKVNDLKKFLNEILGIRDEGKFLIEYKRNHNMWYLD 218

QY 81 -QVCPDGVKCDGKIKYTHKSDNRDVR--NNEDYKPPMGVKPTNITVLYSNEQG---DI 134

Db 219 ERAC-----KYLNSFDDYKNIILGKDMWRDPSIKTENI---LKNFEGIKANI 265

QY 135 TQKLENFCNS-----TNYKDN-NQWEC-----YKDNENRCKLEQNTNNDNPKII 184

Db 266 VSMPSYADLSLDEPRKHWDQNKQKQWAEISCFYKGNHTGVCILMED---DNDN---317

QY 185 SPHNFELWYLLRDTIKWCKLTKTCINNTTHCIDE-----CNRNC 227

Db 318 QYLHFWREWKNDFCIDKLNWDVIK-----EPICDKKVKSPSPNSPDVATVCKNSC 370

QY 228 LCFDRWVQKEEWSIKKLFTHKKNIOQSYNS-INNL 265

Db 371 TDYDKWILNKREY---KMQSSKYKRDRLFNVIQNL 405

RESULT 6

US-09-924-154-15

; Sequence 15, Application US/09924154

; Patent No. US2002012741A1

; GENERAL INFORMATION:

; APPLICANT: Narum, David L.

; APPLICANT: Sim, Kim L.

; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use

; FILE REFERENCE: 05213-0465 43170-262105

; CURRENT APPLICATION NUMBER: US/09/924,154

; CURRENT FILING DATE: 2001-08-07

; PRIOR APPLICATION NUMBER: US 60/223,525

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 15

; LENGTH: 1086

; TYPE: PRT

; ORGANISM: Mammalian

US-09-924-154-15

Query Match

Best Local Similarity 18.4%; Pred. No. 0.003;

Matches 87; Conservative 64; Mismatches 146; Indels 175; Gaps 20;

QY 8 YLSNDKFNINSEYKOFYKELKETOYATNDTF--LNLNEG---KYCKGGLPGEKDI 62

Db 212 YFAKDEPQIVRWLEWSKQFLDEKNYMLFTLRNTYNEMNIIHENCKOYNKVVQNRKKEW 271

QY 63 TFTNSADKK-----GIF--YRS-----EYQVCPDGVKCDGKVT 96

Db 272 TFLSNEFNKIFPERNVQIHSINIFKEYKENVNVDIIFGLTNYEYNNFCKEKLPELVSAAYN 331

QY 97 HKSDNDRE-RV-----NNEDYKPPMGVKPTN-----121

Db 332 LKAPNAKSPRIYKSKHEESSVFGCKTKISKVKKWCYNSNKKVTKPEGVCGPPRRQQLC 391

QY 122 ---ITVLYSGNEQG-----DITQKLEN-----FCNST-NYKD-----150

Db 392 LGYIFLIRDCNEEGLKDHINKAANYEAMHLKEKENAGDKICNATILGSYADIGDIVRGL 451

QY 151 -----KKNQKWCYKVDENINRCKLEQNTENN-----DNPCKII 184

Db 452 DVWRDINTNKLSEKQKIFMGSGNSRKQKQNDNNERNKWKQKQRLINSSVVKHPIPKKTC 511

QY 185 SFHNFELWYLLRDTIKWCKL-----KTCINNTTHCIDEENRCLCFDR 232

Db 512 KRHNFEK-IPQFLRWLKEWGECEMGTEVQKLEKICENKNCSE--KKCKNACSSYEK 568

QY 233 VWKQKEEWSIKKLFTHKKNIOQSYNSINNFEGYFFKVMKLDKDEAKWELMENIK 292

Db 569 WIKERKNEYNLQSKKFSKDKLNKK-----NNLYN-----KFEDSKAYLR 608

QY 293 RKKNEFSNLENNRDYLENAIELLDLHLKETATICKDNNTNEACETSHNATN 344

Db 609 SESQCSNIEFN-----DETFTF--PNKYEACMVCEPNPSS 643

RESULT 7

US-09-924-154-13

; Sequence 13, Application US/09924154

; Patent No. US2002012741A1

; GENERAL INFORMATION:

; APPLICANT: Narum, David L.

; APPLICANT: Sim, Kim L.

; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use

; FILE REFERENCE: 05213-0465 43170-262105

; CURRENT APPLICATION NUMBER: US/09/924,154

; CURRENT FILING DATE: 2001-08-07

; PRIOR APPLICATION NUMBER: US 60/223,525

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 13

; LENGTH: 1421

; TYPE: PRT

; ORGANISM: Mammalian

US-09-924-154-13

Query Match 7.7%; Score 147; DB 10; Length 1421;

Best Local Similarity 23.7%; Pred. No. 0.0089;

Matches 68; Conservative 45; Mismatches 118; Indels 56; Gaps 16;

QY 94 KYTHKSDNRERVANNEDYKPPMGVKPTNITVLYSNEQ-GDIT-QKLENFCNSSTNYKDK 151

Db 530 KYKNDKDEVCIIKNTF-----ADIRDIIGTGYNDLSNRKLVGKINTNSYVHR 581

QY 152 NNQK-----WECYKDE-NINRCKLEQNTNNDN-PKILSFHNFELWYLLRD- 200

Db 582 NKQNDKLFREDWVKYKDKVNVNVSFWFKTKVCKEDDIENIPQFFRWFSEWGDYCDQK 641

QY 201 -----TIKWNCKLTKTCINNTTHCIDEENRCLCFDRWVKQKEEWSIKKLFTHK-KNI 254

Db 642 TKMIETLKECKEPCED-----DNCKRCNSYKWKISKRKEYNKQAKQOYQYQKGN 694

QY 255 QQSYNSINNLFF-EGYFFKVMKLD-----KDEAKWELMENIKRKN-----EFSNL 301

Db 695 NYKYSEFSKPEVLYKSEKSNLNFEDFK--EELHSDYKKNCTMCPEVKDVPISII 753

QY 302 ENRDYLENAI---ELLDLHLKETATICKDNNTNEACETSHNATNP 345

Db 754 RNNEQTSQEAPEESTEAHTETRT---DERKNO--EPANKDLKNP 795

RESULT 8

US-10-153-273-6

; Sequence 6, Application US/10153273

; Patent No. US20020169305A1

; GENERAL INFORMATION:

; APPLICANT: Sim, Kim L.

; APPLICANT: Chitnis, Chetan

; APPLICANT: Miller, Louis H.

; APPLICANT: Peterson, David S.

; APPLICANT: Su, Xin-zhaun

; APPLICANT: Wellens, Thomas E.

TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES

FILE REFERENCE: Q63915
CURRENT APPLICATION NUMBER: PS/09/820.843A

CURRENT APPLICATION NUMBER: US/05/020,045A
CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn version 3.0
SEQ ID NO 109

EQ ID NO 109
LENGTH: 861
TYPE: PRT

LIFE: FRI
 ORGANISM: Plasmodium falciparum
 FEATURE:
 NAME/KEY: misc_feature
 OTHER INFORMATION: hypothetical

OTHER INFORMATION: hypochelical protein

NAME/REF: MISC-RECORDS
OTHER INFORMATION: g11493994
09-820-843A-109

Query Match 7.58:

Best Local Similarity 19.9%; Pred. No. 0.0083; Mismatches 139; Indels 1

10 SDNKEVNNINSEYKOFYERLKETOTATNDTEFLNLN-----EGKYCKGGGLDEKDIQET 65

[illegible][illegible]

```

66 NSADDKGIFYRSEYCVQPCGVKCD-----GIKYTHKSDNDRERVNN 108
      | : | : | : | : | : | : | : | : | : | : | : | : |

```

155 NNINGKVPKLNDDNCYNLPNTNNLYIDKEGKMHLTGKEHYNAASNEYNNHNNKNTNNYNNN 214

109 EDYKPPGWKPTNITVLYSGNEQGDITQKLENCSS--TNYKDKNNQKWECYKDENI 165

215 S-----YNN-----NFCNNYNDNNYNSNKGNGKY-ERSL 247

166 NRCKLEONT---EINN-----DNPKI-ISFHNFFELWV---TYLLRDTIK 203

248 NYLKEHDMVDYEYNNKGNIRKNDSEKYWDNPPLHYSKKNYDIFTLGDIKKYAKNNEKK 307

```

204 WNDK L K T C I N N T T H C I D E C R N C L - - - - - C F D R V V K Q K - - - - 237
      | : |   : | : : | : | :   | : |

```

308 GNNKYMNNHDNNSNNSNNVLNNNNMNSNNSNNYNNLTKDNDEENTKSNFAKWFKNNNMN 367

238 -EEWNSIKKLFTKK-----KNIQSYYSINNLFEGYFFKVMKLDKDE 281

368 VNENTDIIKYLNNKNSQGHSDGKNNNNNGNNIINNNSNKNKNNIFQG----- 414

282 AKWELMENIKRKKNEFSLENNRDYLENAIELLDHLKETATICKDNTNEACETSHNA 341

```

415 -----NSRNYENVMYNNNNN--NNII-----SNNKNEASFNTDNI 449

```

342 TTN 344

450 NTN 452

ULT 10
10-153-273-4

sequence 4, Application US/10153273
 Patent No. US20020169305A1

GENERAL INFORMATION:
APPLICANT: Sim, Kim L.

Chitnis, Chetan
Miller, Louis H.

Peterson, David S.
Su, Xin-zhaun

Wellem's, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX

AND PLASMODIUM FALCIPARUM ERYTHROCYTE BINDING
NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe Martens Olson & Bear

STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach

STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/153,273
FILING DATE: 21-May-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/210,288
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Fuller, Michael
REGISTRATION NUMBER: 36,516
REFERENCE/DOCKET NUMBER: NIH121.1FWDV1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1435 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Plasmodium falciparum
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-153-273-4
Query Match 7.1%; Score 137; DB 9; Length 1435;
Best Local Similarity 20.5%; Pred. No. 0.05;
Matches 86; Conservative 57; Mismatches 149; Indels 128; Gaps 19;
QY 1 YEKEIQSYLSNDN-----KRVN-NINSEYKQFYEKLEKQYATNDTFLNLLNEGYCK- 53
Db 75 FNRKWSYGTDPDNIDKNMSLNNHNEFMNNYQSFLSTS-----SLIKQNKYVPI 126
QY 54 GGLPGKDJFTNSADDKGIFRSEYQVCPDQGVKQDGIKTHSKDNDNR-VNEDYK 112
Db 127 NAVVSKILSFLDSRINNG-RNTSSNNEVLSNCRKRGKMGKWDCKKNDRSYIVCPDR 185
QY 113 PPGVVPPTNITVLYSGNEOD-----ITQKLEN-----FCNSSTN----- 147
Db 186 IQLCIVNLSIIRTYKTRMKDHPLEASKKESQLLLKNDKNKYNKFCNDLKNSFLDYGHL 245
QY 148 -----YKDKNNOK-----WECYKD--E 163
Db 246 AMGNDMDFGYSTKAENKIQVFKGARGHSEHKIKNFRKWNNEPREKLWEAMLSEHKN 305
QY 164 NINRCKLEQNTNNDNPKLIISHPNFELWVYLLRDTIKWNDKLTCTINNTH-CIDE 222
Db 306 NINCKNIPOEEL-QITOWIKWHGEF-----LLERDNRSLKPKSK-CKNNTLYEACEKE 358
QY 223 CNRNCLEFDRWQKKEEWSIKKLTFTKKKNIQOQSYYSINNLFYGFYKVMKDLDKDEA 282
Db 359 CIDPCMRYDRIIRSKFEPWTLSEKVTQKVPKEN-----AENYLIKISE--NKDA 408
QY 283 KWKELMENIKRKNFTSNLNNRDYLENATJELLDHLKETATICK-----DNNTWEACE 336
Db 409 KVSLLNNCAEYSKYCD-----CKHTTTLVKSVLNGNDNTIKERE 450
RESULT 11
US-09-402-100-4
; Sequence 4, Application US/09402100
; Patent No. US20010019834A1
; GENERAL INFORMATION:
; APPLICANT: Daewoong Pharmaceutical Co, LTD
; APPLICANT: Kim, Byung-O
; APPLICANT: Shin, Sung-Seup
; APPLICANT: Yu, Young-Hyo
; APPLICANT: Park, Myung-Hwan
; APPLICANT: Choi, Deok-Joon
; TITLE OF INVENTION: Recombinant Microorganisms Expressing Antigenic Proteins of Hel
; FILE REFERENCE: 0136/0G140
; CURRENT APPLICATION NUMBER: US/09/402,100
; CURRENT FILING DATE: 1999-09-27
; EARLIER APPLICATION NUMBER: KR 97-11950
; EARLIER FILING DATE: 1997-03-31
; EARLIER APPLICATION NUMBER: KR 97-11951
; EARLIER FILING DATE: 1997-03-31
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 4
; LENGTH: 1338
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: CagA/CTXA2B Chimeric protein
; US-09-402-100-4
Query Match 7.1%; Score 136; DB 10; Length 1338;
Best Local Similarity 19.0%; Pred. No. 0.055;
Matches 84; Conservative 69; Mismatches 139; Indels 150; Gaps 17;
QY 2 EKEIQS-----YLSNDNKFVNINSEYKQFYEKLE----- 33
Db 390 QEEIQNKIDFMEFLAQNNAKLDNLSEKEKEKFEITEKDFQKDSKAYLDALGNDRIAFVSK 449
QY 34 -----TQYATNDTFLNLLNEGYCKGGLPGEKDIITNSADDKGIFRSEYQVCP 84
Db 450 KDRHSALITEFGNGDLSYTLKDYKKADRALDREKNVTLQGSGLKHGVMF----- 500
QY 85 DCGYKCDGIKTHKSDNDRE-----RVNNE 109
Db 501 ---VDYSNFKYTNASKNPNKGVGTNGVSHLEVGFKVAIFNLPDLNLAITSFVRNLE 557
QY 110 DYKPPWGVKPTNITVL-----YSGNEQGDITQKLENF-----CNSSTNYK-----D 150
Db 558 DKLTITGLSPQEAANKLIKDFLSSNK--ELVGKTLNFKAVADAANTGNYDEVKKAQKDL 615
QY 151 KNNQWECYKD-----ENINRCKLEQNTNNDNPKLIISHPNFELWVYLLRD--T 201
Db 616 KSLRRKREHLEKEVEKLEKSKGSKNKNKMEAKAQAQNSOKDEIFALIN-----KEANRDARA 669
QY 202 IKWNDKLTCTINNTHCIDECNRNCLCFDRWVKQKEEWSIKKLTFTKKKNIQOQSYSN 261
Db 670 IATAQNLKG-IRKELSDKLENNVKNLKD-----KSFDEFKNGKNKDSKAEETLKALKGS 725
QY 262 INNLFEGYFFKVMOKLDKDEAKWELMENIKRKNFTSNLNNRDY-----LENAT 312
Db 726 VKDL-----GINPEWISKVENNAALNEFKN-GKNKDFSKVTQAKSDLENSV 771
QY 313 ELLDLHLEKATETICKDNNTNEA 334
Db 772 KDVIINQKVTD---KVDNLNOA 790
RESULT 12
US-09-924-154-17
; Sequence 17, Application US/09924154
; Patent No. US20020127241A1
; GENERAL INFORMATION:
; APPLICANT: Narum, David L.
; APPLICANT: Sim, Kim L.
; TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
; FILE REFERENCE: 05213-0465 43170-262105
; CURRENT APPLICATION NUMBER: US/09/924,154
; CURRENT FILING DATE: 2001-08-07

PRIOR APPLICATION NUMBER: US 60/223,525
 PRIOR FILING DATE: 2000-08-07
 NUMBER OF SEQ ID NOS: 17
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 17
 LENGTH: 1501
 TYPE: PRT
 ORGANISM: Mammalian
 US-09-924-154-17

Query Match 6.9%; Score 133; DB 10; Length 1501;
 Best Local Similarity 22.1%; Pred. No. 0.1;
 Matches 79; Conservative 46; Mismatches 153; Indels 80; Gaps 15;

QY 7 SYLSNDKFNINSEYKQFYKLEK---TOYATNDTFLNLLNEGYCKGGPLGKEDIT 63
 DB 474 NYLLDDIY-NVHNSQLLIBIIMASKOEGKLLWKHGTILDONACKYIINDSYVDYKDI 532
 QY 64 FTSADDKGFYRSEYQVCPDGVKCDGKIKYTHKSDNDRERNEDYKPPWGVKPTNIT 123
 533 IGNDL-----WNNNSIKVON-----NLN 551
 QY 124 VLXSGNEQGDITQKLENFCSNSTN--YKDNKQWE---CYKNDENINRCKLEQNTPEIN 177
 DB 552 LIPERFYGKVGKRLFKTKELKNVWILNRKNVWESMRGIDEDVDRKTCERIDELE 611
 QY 178 NDNPKIISFINFELWYLLRTDIKNDKL--KTCINNTTTHCIDB-CNRNCLCFDRW- 233
 DB 612 N---MPQFFRFSQWAHFFCKEYKELNDKCTGNGKSLCQDKTCQNVCTNMNYT 667
 QY 234 -VKOKEEWSIK-----KLFT--KKKNIQSYVSNINNLFFEGYFFKVMKLDKDEAKWK 285
 DB 668 YTRKLAIEIOSYKDKRKLFLSLAKRNVTFLEKAKNSINDFTKIFDQDK---LFK 724
 QY 286 ELMENIKRKNFENLNDRDYLENATIELLDHLKETATICKDNNTNEACETSHNAT 343
 DB 725 ERCSMDTQVLEVNKE-----MLSDNSSEADTIDSEKNGEELYVNHNSVS 772

RESULT 13
 US-09-924-154-14
 Sequence 14, Application US/09924154
 Patent No. US20020127241A1
 GENERAL INFORMATION:
 APPLICANT: Narum, David L.
 APPLICANT: Sim, Kim L.
 TITLE OF INVENTION: Anti-Plasmodium Compositions and Methods of Use
 FILE REFERENCE: 05213-0465 43170-262105
 CURRENT APPLICATION NUMBER: US/09/924,154
 CURRENT FILING DATE: 2001-08-07
 PRIOR APPLICATION NUMBER: US 60/223,525
 PRIOR FILING DATE: 2000-08-07
 NUMBER OF SEQ ID NOS: 17
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 14
 LENGTH: 1143
 TYPE: PRT
 ORGANISM: Mammalian
 US-09-924-154-14

Query Match 6.7%; Score 128; DB 10; Length 1143;
 Best Local Similarity 19.2%; Pred. No. 0.18;
 Matches 84; Conservative 70; Mismatches 151; Indels 132; Gaps 21;

QY 3 KEIQSYLSDN-----KFNVINSEYKQFYKLEKQYATNDTFLNLLNEGYCKGG 55
 DB 74 KLINSFVENSVKKRSLSFINNKTSS-----YDIIPPSYSRNDKF-NSLSENEEDNSGN 127
 QY 56 LPGEKDTFTNSA-----DDKGFYRSEYQVCPDGVKCDGKIKYTHKSDNDRERY- 106
 DB 128 TNSN---NFANTSISIGKDNKQYTFQKRTHLFA-CGIRKRSIKWICRENSKITVCP 183
 QY 107 -----NNEYKPPWGVK-PTNTIVLYSGNEQGD-----ITQKLENF 141

DB 184 DRKIQLCIANFLNSRLETMEKFRIFLISVNTKALLYNKNEGKDSIFCNEIARNSFSDP 243
 QY 142 CNS-----STNYKDNKQWEYKYKDENI-----NRCKLEQNTPEIN 177
 DB 244 RNSFIGDDMDFGGNTDRVKGYINKKFSYKKEKNVEKNLNKKEWEKKNANLWNNHIVN 303
 QY 178 N-----DNPKIISFHFNFELWYLLRTDIKNDKL-----KTCI 212
 DB 304 HGNISKECAIIPAEPOI-----NLWIK-----EWNEFLMEKKRLFLNKDKCV 349
 QY 213 NNTTTH-CIDECNRNCLCFDRWVKQKEEWSIKKLTFTKKKNIQSYVSNINNLFFEGYFF 271
 DB 350 ENKYEACFCGCRPLPCSSYTSFPMKSKTQMEVLNLY-KKNSGVDKNNFLNDFKKNK 408
 QY 272 KVMDKLDKDEAKWELME-----NIKRNKNEFSNENRDRDYLENATIELLDHLKETATI 325
 DB 409 NLDLDFEKEKEYDDLDCCRYTATIK-----SPLNGPAKND-VDIASQINVDLGRFGCN 463
 QY 326 CKDNNTNEACETSHNAT 342
 DB 464 YKSNN-----EKSWNCT 475

RESULT 14
 US-09-820-843A-73
 Sequence 73, Application US/09820843A
 Publication No. US20030039963A1
 GENERAL INFORMATION:
 APPLICANT: Council of Scientific and Industrial Research
 TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PRO
 FILE REFERENCE: 063915
 CURRENT APPLICATION NUMBER: US/09/820,843A
 CURRENT FILING DATE: 2001-03-30
 NUMBER OF SEQ ID NOS: 118
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 73
 LENGTH: 1805
 TYPE: PRT
 ORGANISM: M. genitalium
 FEATURE:
 NAME/KEY: misc feature
 OTHER INFORMATION: gi|1045905
 US-09-820-843A-73

Query Match 6.4%; Score 122; DB 9; Length 1805;
 Best Local Similarity 20.5%; Pred. No. 0.85;
 Matches 81; Conservative 57; Mismatches 121; Indels 136; Gaps 18;
 QY 6 QSYLSNDKFNINSEYKQFYKLEKQYATNDTFLNLLNEGYCKGGPLGKEDITFT 65
 DB 1234 QTYLANKN-----AEY-----SQOQLOQKYTNLLD-----LKENLERTKD- 1269
 QY 66 NSADDKGIFYRSEYCOV---CPDCGVKCDGKIKYTHKSDNDRERNEDYKPPWGVKPTNI 122
 DB 1270 -QDKK---HRSIFARLTKFANDRFEKKQLLKAQRIYVDDKNRLKENER----- 1315
 QY 123 TVLYSGNEQGDITQKLENFCSNSTNYKDNKQK-----WECYKNDENINRCKLE-- 171
 DB 1316 NLHLSNE---TERKRAVLEQDQISYFEKQKQATDAILASHKEVKKKEGELQKLLVELE 1371
 QY 172 -QNTENNDNPKIISFINFELWYLLRTDIKNDKLK-----TCINNTTTHC 219
 DB 1372 TRKTKLNDFAKFSRQREEF-----NORLKLLEQKLTQTOTNSNFKTKA 1418
 QY 220 IDECNRLCLCFDRWVKQKEEWSIKKLTFTKKKNIQSYVSNINNLFFEGYFFKVDKLDK 279
 DB 1419 IQEITENS-----YKRGMEELNFKQKEFDKN-----SRLYE-YFRMRDEIER 1460
 QY 280 DEAKWKELMENIKRKNFES-----NLENNR-DYLENAIELLDHL----- 319
 DB 1461 KESQVKLVKLTQKRNALLEAQANKLNIEKNTIDFKELAKFDKDDIDSTNKORKE 1520

QY 320 -----KETATICKDNNNEACET 337
Db 1521 LNELNENKLLQSLIERERAINSKDLSLLNKKIET 1555

RESULT 15
US-09-286-488-25
; Sequence 25, Application US/09286488
; Patent No. US20020169136A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Sleath, Paul R.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DIAGNOSIS
; TITLE OF INVENTION: AND TREATMENT OF B. MICROTI INFECTION
; FILE REFERENCE: 210121.426C3
; CURRENT APPLICATION NUMBER: US/09/286,488
; CURRENT FILING DATE: 1999-04-05
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 25
; LENGTH: 463
; TYPE: PRT
; ORGANISM: Babesia microti
US-09-286-488-25

Query Match 6.3%; Score 120; DB 9; Length 463;
Best Local Similarity 22.4%; Pred. NO. 0.24;
Matches 92; Conservative 43; Mismatches 146; Indels 130; Gaps 20;

QY 2 EKEIQSYLSNDNKFVNNINSEYKQYKELKETQYATNDTEL-----NLLNE 48
Db 84 DRSQDQKLSHPNKKIKIKISDIIEFDDNAKLPTGSVNDISITCKHNNPVLIRFSLIE 143
QY 49 GKCYK-----GGLPGEKIDITFTNSADKGI-FYRSEYQVQVPCD--GVKCDGI 93
Db 144 GSCYFYFLLNNDTNKNNHKLKYDKTYNEHTDNNNGINYYKIDYSESTEPTTCTCF 203
QY 94 -KYTHKSDNDRRVNNEDYKPPWGVKPTNITVLYSGNEQDITQKLEFCN-----SS 145
Db 204 RKNHKS-----ERKELENYK-----YEGTELARI-----HCNKGKVKLGD 240
QY 146 TNYKDNKNQKWCYKDEINRCKLEQNTNNDNPKIISFHFELWYTLIRDTIKWN 205
Db 241 IKIKDN--LEIYVKQ-----LMSVNPVNFNDPTSIN-----LPTVSTTNDIT-N 284
QY 206 DLKTCIN-NTTTHC-----IDECNRNCLC-----FDRWV 234
Db 285 KYGTIINANIVEYCEFEDEPLTIGFRTIDKSSQNKLSHPNKKIDKIKFDFYIIEFDDV 344
QY 235 KOREEENSIKLETKKN-----IQOSYV-----SNINLFEYFPKVMKLD 278
Db 345 KLPTIGTVNIYYITCBHNPVLFVIEVSIESEYFYFYSMMNNTKNNHKLKYDKRFK 404
QY 279 KDEAKWELMENIKRK-----KNFNSLNENNRDYL--ENAIELLDHLKE 321
Db 405 KYTKNGINCYEYVLRKCSSYTRKNVEYHKLARIHCNEEKCVNVKVDNIEK 455

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:33:00 ; Search time 58.8889 Seconds
(without alignments)
780.648 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 345

Sequence: 1 YEKEIQSYLSNDNKFVNNIN.....CKDNNNEACETSHNATNP 345

Scoring table:

OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 908470 seqs, 133250620 residues

d size: 9

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: A_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	345	100.0	3542	22	AAB62142 P. falciparum FCR3

ALIGNMENTS

RESULT 1:
AAB62142
ID AAB62142 standard; Protein; 3542 AA.
XX

AC AAB62142;

XX 29-MAY-2001 (first entry)

XX P. falciparum FCR3.varCSA protein.

XX FCR3.varCSA protein; chondroitin sulfate A; CSA; var gene: PREMP1;
KW erythrocyte membrane protein 1; parasitized red blood cell; PRBC;
KW malaria; protozoacide.

XX Plasmodium falciparum.

XX WO200116326-A2.

XX 08-MAR-2001.

XX 01-SEP-2000; 2000WO-US24195.

XX 01-SEP-1999; 99US-0152023.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Scherf A, Miller LH, Gamain B, Baruch DI, Buffet P, Scheidig C;
PI Gysin J, Pouvelle B, Fujii N, Smith J;

XX WPI; 2001-235109/24.

DR N-PSDB; AAF57301.

XX Novel FCR3.varCSA protein, useful for modulating parasitized red blood cell binding, sequestration and onset of maternal malaria -

XX Claim 12; Page 63-71; 78pp; English.

XX The invention relates to a P. falciparum FCR3.varCSA protein, that is capable of binding to chondroitin sulfate A (CSA). The var gene and the corresponding P.falciparum erythrocyte membrane protein 1 (PEMP1) modulate adhesion of parasitized red blood cell (PRBC) to CSA. The protein and the encoding gene are useful for treating and preventing maternal malaria in a patient identified at a risk for contracting maternal malaria or in a patient afflicted with maternal malaria. The present sequence represents the P. falciparum FCR3.varCSA protein.

XX Sequence 3542 AA;

Query Match 100.0%; Score 345; DB 22; Length 3542;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YEKEIQSYLSNDNKFVNNINSEYKQFYKLTQYATNDTFLNLLNEGYCKGLPGEK 60

Db 403 YEKEIQSYLSNDNKFVNNINSEYKQFYKLTQYATNDTFLNLLNEGYCKGLPGEK 462

QY 61 DIFTTSADKGIYFIRSEYQVCPDCGVKCDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 120

Db 463 DIFTTSADKGIYFIRSEYQVCPDCGVKCDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 522

QY 121 NITVLYSGNEQGDITQKLENFCNSSTNYKDNKQKWEYKDKENINRCKLEQNTENNND 180

Db 523 NITVLYSGNEQGDITQKLENFCNSSTNYKDNKQKWEYKDKENINRCKLEQNTENNND 582

QY 181 PKLIISHNFFELWVYLLRDTIKWNDKLTICINNTTHCIDEKNRNCICDFRWVKQEEE 240

Db 583 PKLIISHNFFELWVYLLRDTIKWNDKLTICINNTTHCIDEKNRNCICDFRWVKQEEE 642

QY 241 WNSIKLFTKKKNIQSYYSNINNLFGYFFKVMKDDEAKWKELMENIKKKNEFSN 300

Db 643 WNSIKLFTKKKNIQSYYSNINNLFGYFFKVMKDDEAKWKELMENIKKKNEFSN 702

QY 301 LENNRDYLENAIELLDHLKETATICKDNNNTNEACETSHNATNP 345

Db 703 LENNRDYLENAIELLDHLKETATICKDNNNTNEACETSHNATNP 747

Mon Apr 28 13:47:48 2003

us-10-087-013-2_copy_403_747.oligo9.rag

Page 2

Search completed: April 28, 2003, 10:49:42
Job time : 59.3889 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:48:21 ; Search time 31.1111 Seconds
(without alignments)
1066.062 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 345
Sequence: 1 YEKEIQSYLSNDNKFVNIN.....CKDNTNEACETSHNATTNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283224 seqs, 96134422 residues

d size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description

No matches found

rch completed: April 28, 2003, 10:52:27
time : 31.1111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:35:40 ; Search time 17.7778 Seconds
(without alignments)
804.900 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 345
Sequence: 1 YEKEIQSYLSNDNKFVNIN.....CKDNTNEACETSHNATTNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0
Searched: 112892 seqs, 41476328 residues

d size : 9
Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description

No matches found

Search completed: April 28, 2003, 10:50:21
Job time : 18.7778 secs

THIS PAGE BLANK (USP)

GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:45:26 ; Search time 31.6667 Seconds
(without alignments)
2244.829 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747

Perfect score: 345

Sequence: 1 YEKEIQSYLSNDNKFVNIN.....CKDNNNEACETSHNATNP 345

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 671580 seqs, 206047115 residues

d size : 9

Total number of hits satisfying chosen parameters: 14

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: SP_TREMBL_21.*
- 2: SP_Archaea.*
- 3: SP_Fungi.*
- 4: SP_Human.*
- 5: SP_Invertebrate.*
- 6: SP_Mammal.*
- 7: SP_Mhc.*
- 8: SP_Organelle.*
- 9: SP_Phage.*
- 10: SP_Plant.*
- 11: SP_Rodent.*
- 12: SP_Virus.*
- 13: SP_Vertebrate.*
- 14: SP_Unclassified.*
- 15: SP_RVirus.*
- 16: SP_Bacteriap.*
- 17: SP_Archeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	345	100.0	3542	5	Q9U5M2	Q9U5M2 plasmodium
2	240	69.6	461	5	Q8T6K3	Q8T6K3 plasmodium
3	240	69.6	465	5	Q8T6K4	Q8T6K4 plasmodium
4	240	69.6	492	5	Q8T6K9	Q8T6K9 plasmodium
5	238	69.0	494	5	Q8T6K7	Q8T6K7 plasmodium
6	233	67.5	510	5	Q8T6L0	Q8T6L0 plasmodium
7	67	19.4	492	5	Q8T6K6	Q8T6K6 plasmodium
8	67	19.4	1615	5	Q8T325	Q8T325 plasmodium
9	49	14.2	3287	5	Q8T326	Q8T326 plasmodium
10	41	11.9	460	5	Q8T6K1	Q8T6K1 plasmodium
11	41	11.9	494	5	Q8T6K0	Q8T6K0 plasmodium
12	39	11.3	427	5	Q8T6K2	Q8T6K2 plasmodium
13	39	11.3	496	5	Q8T6K8	Q8T6K8 plasmodium
14	34	9.9	494	5	Q8T6K5	Q8T6K5 plasmodium

ALIGNMENTS

RESULT 1

Q9U5M2
ID Q9U5M2 PRELIMINARY; PRT; 3542 AA.
AC Q9U5M2
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE FCR3 CSA ligand (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE=20006305; PubMed=10535993;
RA Buffet P., Gamain B., Scheidig C., Baruch B., Oishi S., Fujii N.,
RA Fusai T., Parzy D., Miller L.H., Gysin J., Scherf A.;
RT "Plasmodium falciparum domain mediating adhesion to chondroitin
RT sulfate A: A receptor for human placental infection.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:12743-12748(1999).
DR EMBL; AJ133811; CAB59840.1; -;
DR InterPro; IPR001219; Neurotoxin.
DR InterPro; IPR004258; PFEMP.
DR Pfam; PF03011; PFEMP; 1.
DR PRINTS; PR00284; TOXIN.
FT NON_TER 3542 3542
SQ SEQUENCE 3542 AA; 413089 MW; 970D85E88DA2EC2 CRC64;
Query Match 100.0%; Score 345; DB 5; Length 3542;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 345; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1 YEKEIQSYLSNDNKFVNINSEYKQFYEKLKETATNDTFLNLLNEGYCKGGLPGEK 60
|||||
403 YEKEIQSYLSNDNKFVNINSEYKQFYEKLKETATNDTFLNLLNEGYCKGGLPGEK 462
QY 61 DITFTNSADDKGFYRSEYQVCPDGVKCDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 120
|||||
463 DITFTNSADDKGFYRSEYQVCPDGVKCDGKIKYTHKSDNDRVNNEDYKPPWGVKPT 522
QY 121 NITVLYSGNEQGDITQKLENFCNSSTNYKDKNNQKWEYKDKENINRCKLEQNTENNND 180
|||||
523 NITVLYSGNEQGDITQKLENFCNSSTNYKDKNNQKWEYKDKENINRCKLEQNTENNND 582
QY 181 PKIISPHNFELWVTVLLRDITKWDKLYTCINNTTTHCIDECNRNCLCFDRVWVKQEEE 240
|||||
583 PKIISPHNFELWVTVLLRDITKWDKLYTCINNTTTHCIDECNRNCLCFDRVWVKQEEE 642
QY 241 WNSIKKLFYKKNQIQSYNSINNLFEGYFFKVMYDKLDEAKWELMENIKRKKKEFSN 300
|||||
643 WNSIKKLFYKKNQIQSYNSINNLFEGYFFKVMYDKLDEAKWELMENIKRKKKEFSN 702
QY 301 LENNRDYLENAIELLDHLKETATICKDNNNEACETSHNATNP 345
|||||
703 LENNRDYLENAIELLDHLKETATICKDNNNEACETSHNATNP 747
RESULT 2
Q8T6K3
ID Q8T6K3 PRELIMINARY; PRT; 461 AA.
AC Q8T6K3;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;

```

[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-M58;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 0:0-0(2002).";
DR EMBL; AF458893; AAL84276.1; -.
FT NON_TER 1 461
FT SEQUENCE 461 AA; 54946 MW; F4C8863D50DA50EF CRC64;
SQ

Query Match 69.6%; Score 240; DB 5; Length 461;
Best Local Similarity 100.0%; Pred. No. 9.7e-237;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 180 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 239
DB 74 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
DB 240 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 299
QY 134 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 300 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 359
QY 194 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 360 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 419

RESULT 3
QY 14 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 180 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 239
DB 74 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
DB 240 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 299
QY 134 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 300 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 359
QY 194 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 360 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 419

[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-M57;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 0:0-0(2002).";
DR EMBL; AF458892; AAL84275.1; -.
FT NON_TER 1 465
FT SEQUENCE 465 AA; 55496 MW; 6B0E35F62468A37B CRC64;
SQ

Query Match 69.6%; Score 240; DB 5; Length 465;
Best Local Similarity 100.0%; Pred. No. 9.8e-237;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 183 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 242
QY 74 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
DB 243 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 302
QY 134 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 303 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 362

[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-M57;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 0:0-0(2002).";
DR EMBL; AF458892; AAL84275.1; -.
FT NON_TER 1 465
FT SEQUENCE 465 AA; 55496 MW; 6B0E35F62468A37B CRC64;
SQ

Query Match 69.6%; Score 240; DB 5; Length 465;
Best Local Similarity 100.0%; Pred. No. 9.8e-237;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 183 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 242
QY 74 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
DB 243 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 302
QY 134 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 303 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 362
```

```

QY 194 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 363 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 422

RESULT 4
QY 194 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 363 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 422

[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-S120;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 0:0-0(2002).";
DR EMBL; AF458887; AAL84270.1; -.
FT NON_TER 1 492
FT SEQUENCE 492 AA; 58590 MW; 277204FDF55014C6 CRC64;
SQ

Query Match 69.6%; Score 240; DB 5; Length 492;
Best Local Similarity 100.0%; Pred. No. 1e-236;
Matches 240; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 73
DB 182 KVVNNINSEYKQFYKELKETQYATNDTFLNLLNEGKYCKGGLPCEKDIITFTNSADDDGI 241
QY 74 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 133
DB 242 FYRSEYQVCPDQGVKCDGKIYTHKSDNDRVNNEDYKPPGWGKPTNITVLYSGNEQGD 301
QY 134 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 193
DB 302 ITQKLEFNCNSSTNYKDKNNQWECYKDEINRCKLEQNTNINNDNPKIISFHNFFELW 361
QY 194 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 253
DB 362 VTYLLRDTIKWNDKLTCTCINNTTTHCHIDECNRNCLCFDRWVKQKEEWNISIKLFTKKKN 421

[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-S120;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
RL J. Infect. Dis. 0:0-0(2002).";
DR EMBL; AF458889; AAL84272.1; -.
FT NON_TER 1 494
FT SEQUENCE 494 AA; 58773 MW; C6D6731157A1C13A CRC64;
SQ
```


Query Match 69.0%; Score 238; DB 5; Length 494;
 Best Local Similarity 100.0%; Pred. No. 1.1e-234;
 Matches 238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGYCKGGLPGKDKITFTNSADDKGI 73
 DB 182 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGYCKGGLPGKDKITFTNSADDKGI 241

QY 74 FYRSEYQVCPGCVKCDGKIYTHKSDNDRVRNNEGYKPPGKVPNTITVLYSGNEQGD 133
 DB 242 FYRSEYQVCPGCVKCDGKIYTHKSDNDRVRNNEGYKPPGKVPNTITVLYSGNEQGD 301

QY 134 ITQKLEFNCSTNYKDKNNQWECYKDNINRCKLEQNTINNDNPKIISFHNFFELW 193
 DB 302 ITQKLEFNCSTNYKDKNNQWECYKDNINRCKLEQNTINNDNPKIISFHNFFELW 361

QY 194 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKKLTFFK 251
 DB 362 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKKLTFFK 419

RESULT 6
 QBT6L0
 ID QBT6L0 PRELIMINARY; PRT; 510 AA.
 AC QBT6L0;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).
 GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 implicated in malaria in pregnancy."
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL; AF458886; AAL84269.1;
 FT NON_TER 1
 FT NON_TER 510
 SQ SEQUENCE 510 AA; 60713 MW; E3CF55DC04601099 CRC64;

Query Match 67.5%; Score 233; DB 5; Length 510;
 Best Local Similarity 100.0%; Pred. No. 1.5e-229;
 Matches 233; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 14 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGYCKGGLPGKDKITFTNSADDKGI 73
 DB 184 KFNINSEYKQFYKELKETOYATNDTFLNLLNEGYCKGGLPGKDKITFTNSADDKGI 243

QY 74 FYRSEYQVCPGCVKCDGKIYTHKSDNDRVRNNEGYKPPGKVPNTITVLYSGNEQGD 133
 DB 244 FYRSEYQVCPGCVKCDGKIYTHKSDNDRVRNNEGYKPPGKVPNTITVLYSGNEQGD 303

QY 134 ITQKLEFNCSTNYKDKNNQWECYKDNINRCKLEQNTINNDNPKIISFHNFFELW 193
 DB 304 ITQKLEFNCSTNYKDKNNQWECYKDNINRCKLEQNTINNDNPKIISFHNFFELW 363

QY 194 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKK 246
 DB 364 VTYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSIKK 416

RESULT 7
 QBT6K6
 ID QBT6K6 PRELIMINARY; PRT; 492 AA.
 AC QBT6K6;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Erythrocyte membrane protein 1 (Fragment).

GN VAR.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 implicated in malaria in pregnancy."
 RL J. Infect. Dis. 0:0-0(2002).
 DR EMBL; AF458886; AAL84269.1;
 FT NON_TER 1
 FT NON_TER 510
 SQ SEQUENCE 510 AA; 60713 MW; E3CF55DC04601099 CRC64;

Query Match 19.4%; Score 67; DB 5; Length 492;
 Best Local Similarity 100.0%; Pred. No. 8e-60;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 244
 DB 354 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 413

QY 245 KKLFTKK 251
 DB 414 KKLFTKK 420

RESULT 8
 QBT325
 ID QBT325 PRELIMINARY; PRT; 1615 AA.
 AC QBT325;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE PEMP1 (Fragment).
 GN TM284VAR3.
 OS Plasmodium falciparum.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5833;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-TM284;
 RX MEDLINE-21927235; PubMed-11930336;
 RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
 RT "Identification of a conserved Plasmodium falciparum var gene
 implicated in malaria in pregnancy."
 RL J. Infect. Dis. 185:1207-1211(2002).
 DR EMBL; AJ420412; CAD20868.1;
 FT NON_TER 1615
 FT NON_TER 1615
 SQ SEQUENCE 1615 AA; 188639 MW; FCAB96C00B86DEA6 CRC64;

Query Match 19.4%; Score 67; DB 5; Length 1615;
 Best Local Similarity 100.0%; Pred. No. 2.1e-59;
 Matches 67; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 244
 DB 577 SFHNFFELWTVYLLRDTIKWNDKLTCTINNTTHCIDECNRNCLCFDRWVKQKEEWSI 636

QY 245 KKLFTKK 251
 DB 637 KKLFTKK 643

RESULT 9
 QBT326
 ID QBT326 PRELIMINARY; PRT; 3287 AA.
 AC QBT326;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

```
DE PfEMP1 (Fragment).
GN TM180VAR2.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM180.
RX MEDLINE=21927235; PubMed=11930336;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 185:1207-1211(2002).
DR EMBL; AJ420411; CAD20867.1; -.
FT NON_TER 3287 3287
SQ SEQUENCE 3287 AA; 383550 MW; 58F8F866FC244536 CRC64;

Query Match 14.2%; Score 49; DB 5; Length 3287;
Best Local Similarity 100.0%; Pred. No. 9.8e-41;
Matches 49; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

297 EFSNLENNRDYLENAIELLDHLKETATICKDNTNEACETSHNATNP 345
|||||
689 EFSNLENNRDYLENAIELLDHLKETATICKDNTNEACETSHNATNP 737
|||||

RESULT 10
Q8T6K1 PRELIMINARY; PRT; 460 AA.
ID Q8T6K1
AC Q8T6K1;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VAN002.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458895; AAL84278.1; -.
FT NON_TER 460 460
SQ SEQUENCE 460 AA; 55157 MW; 06FC87959F5C7B9E CRC64;

Query Match 11.9%; Score 41; DB 5; Length 460;
Best Local Similarity 100.0%; Pred. No. 2.9e-33;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CINNTTHCIDECCNCLCFDRWVKQKEEWSIKKLFYKK 251
|||||
DB 375 CINNTTHCIDECCNCLCFDRWVKQKEEWSIKKLFYKK 415
|||||

RESULT 11
Q8T6K0 PRELIMINARY; PRT; 494 AA.
ID Q8T6K0
AC Q8T6K0;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN-VAN434;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458896; AAL84279.1; -.
FT NON_TER 494 494
SQ SEQUENCE 494 AA; 59240 MW; 739BE121652A367B CRC64;

Query Match 11.9%; Score 41; DB 5; Length 494;
Best Local Similarity 100.0%; Pred. No. 3.1e-33;
Matches 41; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 211 CINNTTHCIDECCNCLCFDRWVKQKEEWSIKKLFYKK 251
|||||
DB 381 CINNTTHCIDECCNCLCFDRWVKQKEEWSIKKLFYKK 421
|||||

RESULT 12
Q8T6K2 PRELIMINARY; PRT; 427 AA.
ID Q8T6K2
AC Q8T6K2;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TM4C8;
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458894; AAL84277.1; -.
FT NON_TER 427 427
SQ SEQUENCE 427 AA; 50969 MW; 6AF7922CB988F20 CRC64;

Query Match 11.3%; Score 39; DB 5; Length 427;
Best Local Similarity 100.0%; Pred. No. 3e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 SFHNFELWVYLLRDTIKWNDKLTCTCINNTTHCIDECC 223
|||||
DB 354 SFHNFELWVYLLRDTIKWNDKLTCTCINNTTHCIDECC 392
|||||

RESULT 13
Q8T6K8 PRELIMINARY; PRT; 496 AA.
ID Q8T6K8
AC Q8T6K8;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Erythrocyte membrane protein 1 (Fragment).
GN VAR.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;
RT "Identification of a conserved Plasmodium falciparum var gene
implicated in malaria in pregnancy.";
RL J. Infect. Dis. 0:0-0(2002).
DR EMBL; AF458888; AAL84271.1; -.
FT NON_TER 496 496
SQ SEQUENCE 496 AA; 59240 MW; 739BE121652A367B CRC64;
```

SQ SEQUENCE 496 AA; 59269 MW; A9569625061F9411 CRC64;

Query Match 11.3%; Score 39; DB 5; Length 496;
Best Local Similarity 100.0%; Pred. No. 3.4e-31;
Matches 39; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 185 SPHNFELWVYLLRDTIKWNDKLTCTINNTTTHCIDEC 223

DB 355 SPHNFELWVYLLRDTIKWNDKLTCTINNTTTHCIDEC 393

RESULT 14

Q8T6K5

ID Q8T6K5 PRELIMINARY; PRT; 494 AA.

AC Q8T6K5;

DT 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Erythrocyte membrane protein 1 (Fragment).

VAR.

Plasmodium falciparum.

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

OX NCBI_TaxID=5833;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K117;

RA Rowe J.A., Kyes S.A., Rogerson S.J., Babiker H.A., Raza A.;

RT "Identification of a conserved Plasmodium falciparum var gene

implicated in malaria in pregnancy.";

RL J. Infect. Dis. 0:0-0(2002).

DR EMBL: AF458891; AAL84274.1;

FT NON_TER 1

FT NON_TER 494

SQ SEQUENCE 494 AA; 58870 MW; 56A34264A9285321 CRC64;

Query Match 9.9%; Score 34; DB 5; Length 494;

Best Local Similarity 100.0%; Pred. No. 4.5e-26;

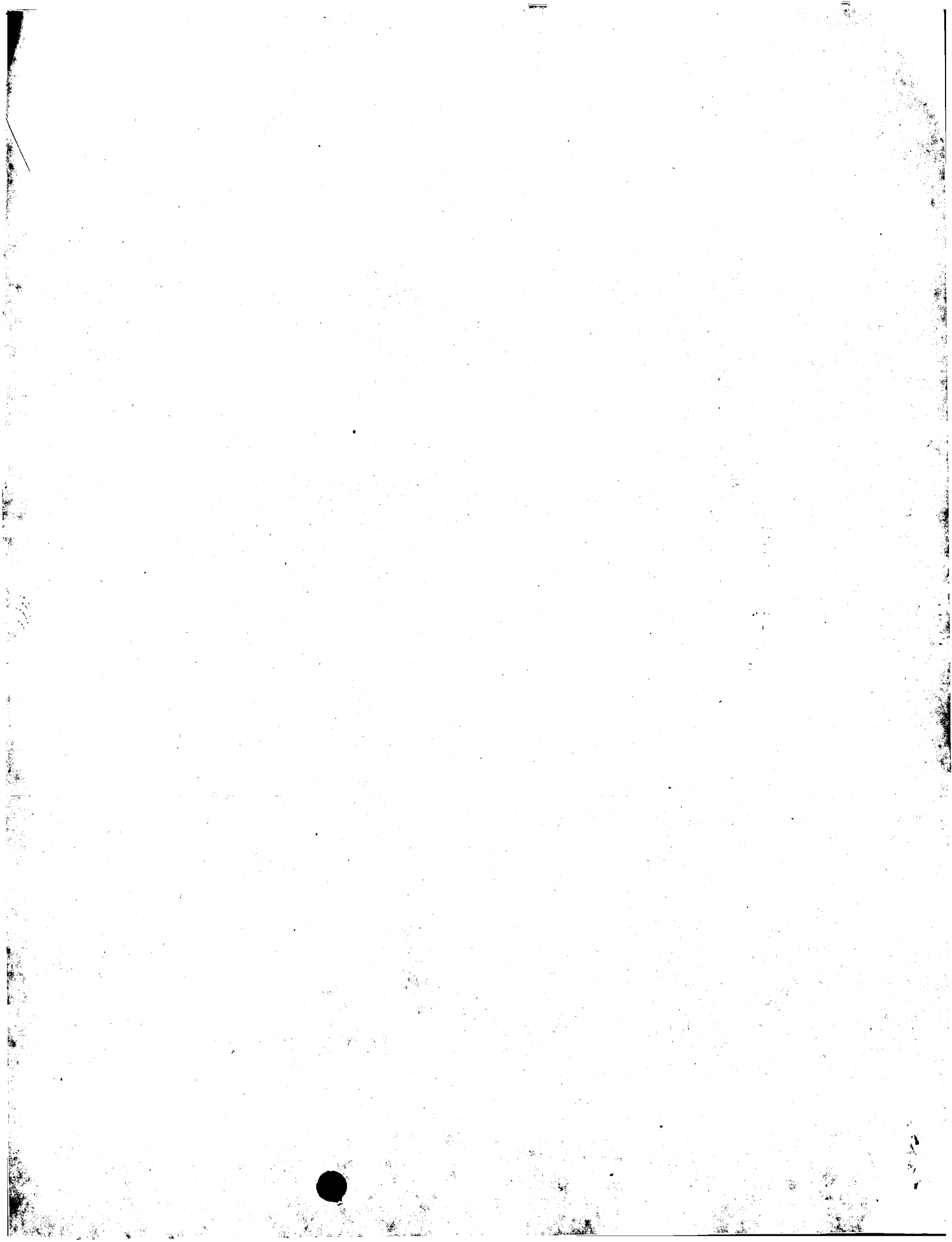
Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 SSTNYKDKNNQWECYKDKENINRCKLEQNTIN 177

DB 315 SSTNYKDKNNQWECYKDKENINRCKLEQNTIN 348

Search completed: April 28, 2003, 10:51:26

Job time : 33.6667 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:49:21 ; Search time 21.1111 Seconds
(without alignments)
480.832 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 345
Sequence: 1 YEKEIQSYLSNDKFNIN.....CKDNTNEACEISHNATNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 262574 seqs, 29422922 residues

d size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Query Score	Match	Length	DB	ID	Description
---------------	----------------	-------	--------	----	----	-------------

No matches found

Search completed: April 28, 2003, 10:53:12
Job time : 21.1111 secs



GenCore version 5.1.4.p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 28, 2003, 10:51:36 ; Search time 16.6667 Seconds
(without alignments)
1658.687 Million cell updates/sec

Title: US-10-087-013-2_COPY_403_747
Perfect score: 345
Sequence: 1 YEKEIQSYLSNDKFNIN.....CKDNTNEACETSHNATNP 345

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 301932 seqs, 80129803 residues

d size : 9

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications AA:
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	DB	ID
Description					

No matches found

Search completed: April 28, 2003, 11:02:43
Job time : 16.6667 secs

THIS PAGE BLANK (USPTO)